

中国科学技术大学

工程硕士学位论文



基于时空大数据的重型货车 NO_x 排放 建模及案例分析

作者姓名： 刘古玥

学科专业： 计算机技术

校内导师： 赵云波 教授

实践导师： 尹航 主任

完成时间： 二〇二五年五月三十日

University of Science and Technology of China

A dissertation for master's degree



Modeling and case analysis of heavy-duty truck NO_x emission based on spatiotemporal big data

Author: Liu Guyue

Speciality: Computer Technology

Supervisor: Prof. Zhao Yunbo

Practice supervisor: Dir. Yin Hang

Completion date: May 30, 2025

中国科学技术大学学位论文原创性声明

本人声明所呈交的学位论文，是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名： 刘古朋

签字日期： 2025年5月10日

中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入《中国学位论文全文数据库》等有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

控阅的学位论文在解密后也遵守此规定。

☒ 公开 ☐ 控阅（____年）

作者签名： 刘古朋

导师签名： 朱武

签字日期： 2025年5月10日

签字日期： 2025年5月23日

摘要

在我国启动新一轮大气污染治理后，移动源防控上升为关键举措。从机动车排放总量看，汽车贡献了 98% 的氮氧化物 (NO_x) 排放，其中重型柴油货车对汽车 NO_x 排放的贡献为 70.75%。由于重型柴油货车具有运行工况复杂等特点，其排放核算存在数据缺失、模型精度不足的困难。因此，不仅需要突破传统核算方法的局限性，更需融合多源数据，以构建准确的重型柴油货车 NO_x 排放清单，掌握其排放特征、实现污染的精准防治。

本研究旨在有效应对重型柴油货车 NO_x 排放精准核算与高分辨率排放清单构建难题。主要难点包括：一是需要获取准确的 NO_x 排放因子以构建宏观尺度排放清单，二是需要获取车辆的瞬态排放数据以建立中微观高分辨率的排放清单。当前重型柴油货车 NO_x 排放因子多基于台架测试或车载尾气检测设备 (PEMS)。台架测试脱离实际路况，PEMS 数据采集量较小，二者均存在滞后性和不确定性。此外，传统排放模型依赖历史数据，忽略车辆的近实时瞬态排放。为解决以上问题，本文基于车载诊断系统 (OBD) 采集真实运行数据，完成了以下主要工作：

(1) 获取了真实道路下的重型柴油货车 NO_x 排放因子。针对当前缺少国六阶段重型柴油货车排放因子的问题，本文利用融合了车辆运行工况和道路特征的时空数据集，获取国六重型柴油货车的 NO_x 排放因子，填补了真实环境下的 NO_x 国六排放因子的空白，分析了道路类型对排放因子的影响。在时空数据集构建过程中，通过线性插值、多层感知机和随机森林实现分阶段插值，修复 OBD 数据缺失。基于时空数据集，得出国六重型柴油货车 NO_x 的基准排放因子为 1.64g/km。另外，快速路 (1.82g/km) 较主干路 (1.28g/km) 和次支路 (1.24g/km) 排放因子高 29.7%-31.9%，揭示路网结构对排放的关键影响。

(2) 构建了重型柴油货车 NO_x 瞬态排放预测模型。针对传统模型忽略近实时瞬态排放的问题，本文基于时空数据，提出了集成轻量级梯度提升机 (LightGBM) 和极端梯度提升算法 (XGBoost) 的 NO_x 瞬态排放预测模型，用于建立高分辨率的 NO_x 排放清单。通过堆叠泛化集成 LightGBM 的高效性和 XGBoost 的抗过拟合能力，深入挖掘时空数据集中 NO_x 瞬态排放与其它八个车辆特征的关系，建立了 SLXGB- NO_x 排放模型，实现对 NO_x 排放的准确预测。与其他方法对比，该模型在效率和预测准确度上都展现出了更佳的表现。利用此模型构建了成都市重型柴油货车 1km×1km 的逐小时 NO_x 排放清单， NO_x 排在时间上呈“双峰”分布 (日间占 79.9%)，并沿绕城高速形成环状聚集带。

关键词：时空大数据；重型货车；排放建模

ABSTRACT

After China launched a new round of air pollution control, mobile source prevention and control became a key initiative. In terms of total motor vehicle emissions, automobiles contribute 98% of nitrogen oxides (NO_x) emissions, with heavy-duty diesel trucks contributing 70.75% to automobile NO_x emissions. Because heavy-duty diesel trucks have complex operating conditions and other characteristics, their emission accounting has the difficulties of missing data and insufficient model accuracy. Therefore, it is necessary not only to break through the limitations of traditional accounting methods but also to integrate data from multiple sources to construct an accurate NO_x emission inventory of heavy-duty diesel trucks, to grasp their emission characteristics, and to achieve accurate prevention and control of pollution.

This study aims to effectively address the difficulties of accurate NO_x emission accounting and high-resolution emission inventory construction for heavy-duty diesel trucks. The main difficulties include: firstly, the need to obtain accurate NO_x emission factors to construct a macro-scale emission inventory, and secondly, the need to obtain transient emission data from the vehicle to build a micro-micro high-resolution emission inventory. Currently, the NO_x emission factors of heavy-duty diesel trucks are mostly based on bench tests or portable emission measurement system (PEMS). The bench test is detached from the actual road conditions, and the PEMS data collection volume is small, both of which have lag and uncertainty. In addition, the traditional emission model relies on historical data and ignores the near-real-time transient emissions of vehicles. To solve the above problems, this study collects real operating data based on the on-board diagnostic (OBD) system, and completes the following main work:

(1) The NO_x emission factors of heavy-duty diesel trucks on real roads were obtained. Because of the current lack of emission factors of heavy-duty diesel trucks in National VI stage, this study uses the spatiotemporal dataset integrating vehicle operating conditions and road characteristics to obtain the NO_x emission factors of National VI heavy-duty diesel trucks, fills in the gaps of the NO_x National VI emission factors under the real environment, and analyses the influence of the road type on the emission factors. During the construction of the spatiotemporal dataset, the missing OBD data are repaired by linear interpolation, multilayer perceptron, and random forest to achieve phased interpolation. Based on the spatiotemporal dataset, the baseline emission factor of NO_x for National VI heavy-duty diesel trucks was obtained to be 1.64 g/km. In ad-

dition, the emission factor of expressway (1.82 g/km) was higher than that of the main road (1.28 g/km) and the sub-branch road (1.24 g/km) by 29.7%-31.9%, which reveals the key influence of the structure of the road network on the emission.

(2) A transient emission prediction model for heavy-duty diesel truck NO_x was constructed. Aiming at the problem that traditional models ignore near-real-time transient emissions, this study proposes a NO_x transient emission prediction model integrating light gradient boosting machine (LightGBM) and eXtreme gradient boosting (XGBoost) based on spatiotemporal data for building a high-resolution NO_x emission inventory. By combining the high efficiency of LightGBM and the overfitting resistance of XGBoost through stacked generalisation, the relationship between NO_x transient emissions and eight other vehicle characteristics in the spatiotemporal dataset is deeply explored, and the SLXGB- NO_x emission model is established to achieve accurate prediction of NO_x emissions. Compared with other methods, the model demonstrated better efficiency and prediction accuracy performance. The model was used to construct an hourly NO_x emission inventory of heavy-duty diesel trucks in Chengdu city for 1km $\times$ 1km, and the NO_x emissions showed a ‘bimodal’ distribution in time (79.9% during daytime) and formed a ring-shaped aggregation zone along the city expressway.

KEY WORDS: Spatiotemporal big data, Heavy-duty trucks, Emission modeling

目录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.2.1 传统排放模型	2
1.2.2 基于 OBD 数据的排放研究	4
1.3 研究目的与内容	6
1.4 本文组织结构	7
第 2 章 相关理论基础	9
2.1 回归算法基础	9
2.1.1 岭回归	9
2.1.2 多层感知机	9
2.2 基于树的集成模型及组合方法	10
2.2.1 随机森林算法	10
2.2.2 极端梯度提升算法	11
2.2.3 轻量级梯度提升机算法	13
2.2.4 堆叠泛化	14
2.3 数据检验与评价指标	15
2.3.1 K-S 检验	15
2.3.2 递归特征消除	16
2.3.3 SHAP 值	16
2.3.4 相关系数	17
2.3.5 均方根误差	17
2.4 本章小结	18
第 3 章 基于时空数据的真实 NO_x 排放因子	19
3.1 OBD 数据采集与处理	19
3.1.1 OBD 数据采集	19
3.1.2 数据预处理	19
3.1.3 分阶段数据插补	22
3.2 工况数据	28
3.2.1 车辆比功率 VSP	28
3.2.2 VSP 区间划分	29

3.3 道路信息数据	30
3.3.1 车辆道路定位	30
3.3.2 道路分类	31
3.3.3 车辆行程划分	32
3.3.4 车辆运行工况分析	33
3.4 NO _x 排放因子	35
3.4.1 NO _x 排放率	36
3.4.2 速度-NO _x 排放因子	37
3.4.3 重型柴油货车 NO _x 排放因子	39
3.5 本章小结	41
第 4 章 基于堆叠泛化的 NO_x 瞬态排放预测方法	42
4.1 数据集与相关性分析	42
4.2 基于 XGBoost 的基础 NO _x 排放模型	43
4.3 基于堆叠泛化的 SLXGB-NO _x 排放预测模型	44
4.3.1 模型建立步骤	44
4.3.2 模型结果分析	46
4.3.3 模型对比	47
4.4 成都市高分辨率 NO _x 排放清单	50
4.4.1 逐小时车辆运行与 NO _x 排放数据	51
4.4.2 NO _x 网格化排放清单	52
4.5 本章小结	54
第 5 章 总结与展望	56
5.1 本文工作总结	56
5.2 未来展望	57
参考文献	59
致谢	65
在读期间取得的科研成果	66

插图和附表清单

图 1.1	重型柴油货车污染物排放量占比	2
图 1.2	本文主要研究内容	7
图 2.1	多层感知机结构图	10
图 2.2	随机森林算法过程图	11
图 2.3	树的生长策略对比	14
图 2.4	Stacking 结构示意图	15
图 3.1	OBD 数据采集终端结构	20
图 3.2	OBD 数据传输过程	20
图 3.3	OBD 数据筛选与清洗流程图	22
图 3.4	各特征缺失占比	22
图 3.5	线性插值测试 K-S 检验	23
图 3.6	K-S 检验填补结果举例	24
图 3.7	Pearson 相关系数热力图	25
图 3.8	Spearman 相关系数热力图	25
图 3.9	MLP-RF 模型简要构建过程	26
图 3.10	MLP-RF 模型特征重要性分析	27
图 3.11	MLP-RF 模型真实-预测值	28
图 3.12	道路信息定位流程图	32
图 3.13	车辆行程划分流程图	33
图 3.14	Motorway 上特定速度区间下 VSP Bin 分布	34
图 3.15	Primary 上特定速度区间下 VSP Bin 分布	35
图 3.16	Secondary 上特定速度区间下 VSP Bin 分布	35
图 3.17	NO _x 排放因子获取过程	36
图 3.18	VSP 区间下的 NO _x 排放数据分布	37
图 3.19	NO _x 排放因子曲线图	39
图 3.20	NO _x 综合排放因子	40
图 4.1	Pearson 和 Spearman 相关系数	42
图 4.2	基础 NO _x 排放模型表现	43
图 4.3	SLXGB-NO _x 排放预测模型	44
图 4.4	SLXGB-NO _x 排放模型特征 SHAP 值	47

图 4.5	SLXGB-NO _x 排放模型效果	48
图 4.6	基础模型对比	50
图 4.7	成都市逐小时车辆数据统计	51
图 4.8	成都市重型柴油货车逐小时 NO _x 排放量	52
图 4.9	成都市主要道路	53
图 4.10	成都市 (a)9 时 (b)14 时里程数据分布	54
图 4.11	成都市 (a)9 时 (b)14 时 NO _x 排放量分布	54
表 1.1	传统排放模型总结	4
表 3.1	OBD 数据集字段说明	21
表 3.2	可用数据集统计	22
表 3.3	RF 参数取值	27
表 3.4	MLP 参数取值	27
表 3.5	模型结果对比	28
表 3.6	VSP 计算公式系数取值	29
表 3.7	重型货车 VSP 区间划分	30
表 3.8	道路分类	32
表 4.1	特征消融对比结果	44
表 4.2	LightGBM 超参数取值范围	45
表 4.3	XGBoost 超参数范围	46
表 4.4	基模型权重	46
表 4.5	模型对比	49

第1章 绪论

1.1 研究背景与意义

随着经济的高速发展，物流运输需求的急剧增长，重型柴油货车作为运输的主要力量，其保有量不断上升。由于重型柴油货车的运行特征以及跨区域流动性，其尾气污染物有直接污染、流动污染的特点。根据生态环境部发布的《中国移动源环境管理年报（2023年）》^[1]可知，2022年，全国机动车氮氧化物（Nitrogen Oxides, NO_x ）排放量为526.7万吨，其中汽车的 NO_x 排放占比超过90%。按照具体的车辆类型与燃料类型来看，2022年重型柴油货车排在汽车总排放中的占比如图1.1，其中重型柴油货车的 NO_x 排放量占汽车排放总量的70.75%。根据国际癌症研究机构发布的《化学物质致癌性评估报告》^[2]可知，柴油发动机尾气被明确列为I类致癌物（对人类致癌）。作为其主要毒性成分之一的 NO_x ，是一种典型的大气污染物，经呼吸道进入人体后可通过多途径产生毒性效应。该气体不仅会通过抑制血红蛋白携氧能力干扰血液供氧功能，其强氧化特性还可引发肺泡上皮细胞损伤，导致肺水肿等急性炎症反应，并通过长期暴露增加呼吸系统肿瘤的发生风险^[3]。由此可知，重型柴油货车作为大气污染物的主要来源之一，其排放的高浓度 NO_x 不仅会对空气质量造成了极大的负面影响，更会对人体健康造成巨大危害。因此，加强对重型柴油货车排放的有效监管，对优化大气质量，改善生活环境，保证人体健康至关重要。

当前对重型柴油货车污染防治的难点和重点集中在对其 NO_x 排放控制方面。这需要依据精准的 NO_x 排放清单，分析其排放特征，以建立重型柴油货车标准控制措施，完善相关政策体系，实现排放管控。不同尺度下精准的排放清单依赖于不同的数据基础，其中建立宏观尺度的排放清单需要以准确的排放因子为基础，中微观高分辨率的排放清单的建立则需要依据车辆的真实运行数据获取其瞬态排放。因此，获取准确的 NO_x 排放因子，以及利用精准的 NO_x 排放模型获取瞬态排放数据是实现重型柴油货车 NO_x 排放控制的基础与核心。

以往用于获取排放因子的数据集多来自于传统的排放测试技术，如实验室台架实验、遥感测试、便携式排放测量系统（Portable Emission Measurement System, PEMS）等方法，存在数据量少、数据精确度低、实验成本高、操作复杂等缺点；用于计算 NO_x 排放数据的传统排放模型未考虑车辆的近实时瞬态排放，逐渐无法满足政府和企业对重型柴油货车排放动态监测、准确评估和管控的要求。近年来，车联网技术的迅速发展、车载诊断系统（On-Board Diagnostics, OBD）的更新以及国家第六阶段机动车污染物排放标准（以下简称国六）^[4]对重型柴油货车OBD数据上传远程终端数据服务器的严格要求，使得采集重型柴油货车实际行

驶过程中的数据成为可能。这不仅降低了车辆数据的获取成本，还提高了数据的规模和真实性。

重型柴油货车作为道路移动源，其流动性强、跨区域运行较为频繁，除了车辆驾驶条件这一主要因素外，车辆的行驶轨迹、外界环境等也会影响车辆的运行状态，进而对其 NO_x 排放造成影响。因此，以 OBD 数据为基础，融合车辆定位、道路信息等多源数据，构建时空数据集，综合考虑车辆自身运行特征与外界因素对 NO_x 排放的影响，建立真实环境下的 NO_x 排放因子和精确的 NO_x 排放模型，为更加高效、精准地获取 NO_x 排放数据，建立高分辨率的 NO_x 排放清单，提供良好的数据基础。这同时对分析重型柴油货车 NO_x 的排放特征，进行排放监管以及相关政策的制订具有重要意义。

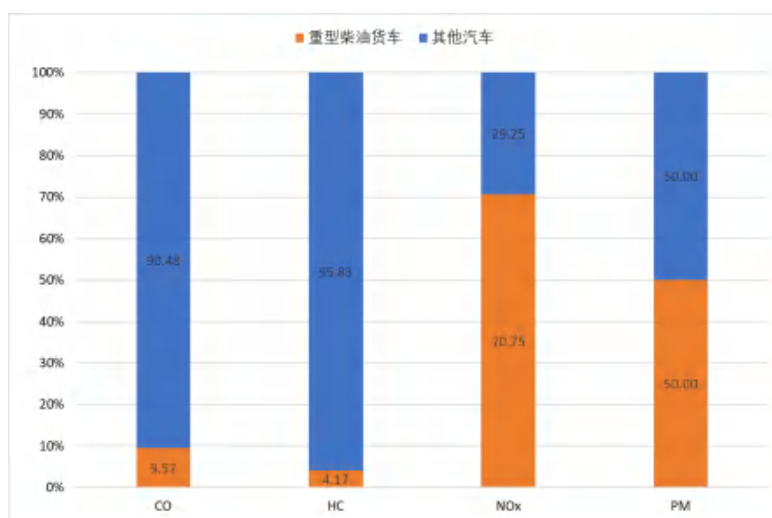


图 1.1 重型柴油货车污染物排放量占比

1.2 国内外研究现状

机动车排放模型通过结合多种影响排放的因素，实现对机动车排放因子的测算，并进一步与交通数据结合，建立区域的排放清单，以此将尾气排放数据量化，实现对机动车排放规律的分析，为排放预测提供支持。这里将排放相关的研究按照数据集的差异分为两类，以进行国内外研究现状阐述：一是基于实验台架测试或者 PEMS 数据的传统排放模型，二是基于 OBD 数据的相关排放研究。

1.2.1 传统排放模型

国内外关于排放模型的研究成果颇丰，目前比较常见并被广泛使用的排放模型总结对比如表1.1所示。基于测试数据集建立的排放模型主要分为宏观与微观两大类。其中，宏观尺度的模型以平均速度为核心变量开展统计回归分析，包括早期由美国环境保护署（Environmental Protection Agency, EPA）开发的机动车排放模拟器（Motor Vehicle Emission Simulator, MOBILE），以及由 Emisia 公司主

导开发, 欧盟委员会联合研究中心提供科学指导、欧洲环境署协调的道路交通排放计算模型 (Computer Programme to Calculate Emissions from Road Transport, COPERT) 等。这种模型对数据的要求较低, 对计算宏观尺度的机动车排放与分析有一定的优势。微观尺度的排放模型主要依托机动车精细化行驶特征数据构建, 常见模型有以车辆比功率 (Vehicle Specific Power, VSP) 作为主要参数、由 EPA 升级 MOBILE 得到的 MOVES 模型、国际清洁交通委员会组织开发的国际机动车排放模型 (International Vehicle Emissions model, IVE), 由多个欧洲国家联合开发、融合平均速度与加速度参数的道路运输排放因子手册 (Handbook of Emission Factors for Road Transport, HBEFA), 以及基于发动机负载等参数的综合模态排放建模 (Comprehensive Modal Emissions Modeling, CMEM) 等^[5-8]。

在以上排放模型的基础上, 国内外学者都利用各种模型展开了相关研究工作。我国已经进行了许多排放模型本地化的相关研究工作, 何巍楠等^[9]以 HBEFA 模型为理论框架, 通过整合北京市本地机动车类型构成、实际运行工况及燃料特性等底层数据, 运用计算机仿真技术构建了适配北京交通特征的温室气体排放模型, 并完成了覆盖车型分类、交通状态、工况单元和活动水平四个维度的本地化交通排放因子库建设。曹杨等^[10]基于深圳实测道路排放数据, 对 MOVES 模型的地理信息、车辆类型以及燃油等微观参数进行本地化校准, 并对比分析了本地化前后 MOVES 模型估算结果与车载实测排放因子的相对误差, 结果表明 MOVES 深圳模型能够有效反映深圳机动车排放特征, 为区域大气污染防治提供了可靠的量化工具。周华等^[11]通过道路实测数据对比分析了 IVE 模型中引擎负载和前平均比功率两种区间划分方法对排放因子估算的影响, 发现前平均比功率方法在一氧化碳、二氧化碳、 NO_x 与比功率的相关性上显著优于引擎负载方案, 且改进后的模型排放因子与实测值更接近, 表明优化区间划分能有效提高 IVE 模型对机动车排放估算的准确性。宋晓伟等^[12]针对长三角城市群机动车污染物排放问题, 基于 COPERT-IV 模型构建了 1999-2017 年的排放清单, 系统分析了 10 种污染物的时空分布特征及车型贡献, 发现轻型客车主导一氧化碳、非甲烷总烃、二氧化碳、甲烷、一氧化二氮和氨排放, 重型货车和客车则是 NO_x 、细颗粒物、可吸入颗粒物和二氧化硫的主要来源, 且污染物排放沿高速公路呈带状分布, 长江下游、杭州湾和太湖周边城市排放强度最高, 为长三角城市群机动车污染治理提供了重要数据支撑。Gao 等^[13]将 IVE 模型与哈尔滨-长春特大城市的道路机动车排放清单技术指南结合, 对车辆排放的时空特征进行分析, 发现柴油重型卡车是 NO_x 和颗粒物的主要来源。

国外的许多研究人员也利用排放模型完成了许多的相关工作, 如 Nagpure 等^[14]将印度本地化的地形、气象等特定参数输入 IVE 模型, 在印度三个高海拔城市中测试了海拔对轻型汽油商用车排放的臭氧前体物 NO_x 、一氧化碳、挥发

性有机物的排放率的影响，结果表明一氧化碳与挥发性有机物的排放率和海拔高度成正比。Wallace 等^[15]使用 MOVES 与 MOBILE6.2 两种模型，分别预测了车辆在冬季各时间段的一氧化碳与 NO_x 的摩尔排放比，经过与实际观测结果对比得到：MOVES 模型的结果更接近实际。Ritner 等^[16]利用 MOVES 模型，计算出不同行驶工况下的车辆排放因子，并用于 CAL3QHC 模型，研究十字路口车辆加速和减速排放因子的预测，结果数据显示：MOVES 模型的加入大大提高了预测的准确度。

这些排放模型虽然在宏观尺度排放清单构建中表现稳健，但其核心算法依赖静态工况假设，且在数据集获取、动态模拟精度以及成本与计算资源等方面存在一定的限制，难以捕捉真实驾驶中的瞬态波动。

表 1.1 传统排放模型总结

分类	模型	研究尺度	优点
基于平均速度	MOBILE	宏观	排放因子数据库拥有大量实测数据。
	COPERT	宏观	研究涉及的污染物种类最多可达28种，该模型可用于计算年排放量。
基于VSP工况	MOVES	微观	以VSP为参数，可以用于道路排放和非道路排放的研究。
	IVE	微观	引入了VSP分布，能够更真实的反映出运行工况和机动车排放之间的关系。
基于加速度	HBEFA	微观	基于平均速度、加速度等行驶特征参数构建排放模型，通过分析车辆行驶过程中的速度变化区间、加速度状态等，结合车辆类型等信息计算排放因子。
基于发动机参数	CMEM	微观	基于微观形式工况具有独到的计算优势。

1.2.2 基于 OBD 数据的排放研究

为了降低数据的获取成本，扩大数据的获取范围，提升数据的有效性，当前利用车载诊断系统（OBD）获取车辆的真实运行数据已成为一大趋势。OBD 设备是最初是为了诊断汽车故障而设计的标准化系统，当前发展到了 OBD-III 阶段。其中，在 OBD-I 阶段，该设备从外部连接到汽车的控制系统，仅能对车辆的部分部件监测。OBD-II 阶段技术逐渐规范化，主要用于车辆故障排查维修和排放监测，但是信息只存储于车辆的控制模块中，需要外接设备才可以读取。随着车载通讯方面技术的进步，当前阶段的 OBD-III 设备具有了监测、维护和车辆信息发送多种功能。该设备利用车载通讯系统将车辆运行信息发送到管理部门的远程监控平台，其中包括运行数据、排放数据、故障代码等，用于车辆排放管理等工作^[17-20]。OBD 数据集在交通流量监测、机动车排放监测、排放清单建立等方面已得到广泛应用。

国内外利用 OBD 设备及其数据集开展了大量研究工作。国内,Zhang 等^[21]利用车辆发动机的 OBD 参数和排放传感器的数据,研发用于中国重型柴油车排放监测的车载监测(OBM)系统,并通过 PEMS 对 8 辆装有 OBM 的重型柴油车进行道路测试,发现 OBM 和 PEMS 结果在多数情况下一致性较好。Jiang 等^[22]基于 OBD 系统能够实时监测车辆发动机和后处理系统状态、及时发现可能影响尾气排放的故障的功能,通过实际测试和案例分析,证明了基于 OBD 的维修在减少 NO_x 和颗粒物排放方面具有显著效果,为实施全面的重型柴油车的检查和维护计划提供了科学依据和实践参考。Wang 等^[23]针对轻型汽油车的城市碳排放难以量化的问题,利用 PEMS 评估黑碳的排放,并结合机器学习模型,利用 OBD 数据预测黑碳排放,在车辆中集成模型,实现对黑碳排放的实时监控,帮助制定减排策略实现。Hao 等^[24]利用在中国重型车商用车测试工况下对某重型载货汽车进行底盘测功机排放测试获取的数据,对基于 OBD 数据计算的车辆 NO_x 和二氧化碳排放进行了验证,结果表明,基于 OBD 数据计算的 NO_x 和二氧化碳排放与排放分析仪实测结果一致,证明了 OBD 数据在排放计算中的可用性。Lv 等^[25]利用质量尚可的 OBD 数据集计算了重型车的瞬时排放,但通过对唐山市境内行驶的重型柴油货车的 OBD 数据进行质量评估发现:超过 24% 的车辆其数据总体失效率超过 20%,且只有 0.4% 的车辆满足 1Hz 的传输周期要求,指出了 OBD 数据急需解决的质量问题。

国外,Fernandes 等^[26]将 PEMS、OBD 以及定位数据结合,获取四辆车辆的实际排放与发动机数据,分析了驾驶行为对排放的影响,并利用柴油车内部可观察变量建立排放预测模型,发现了内部可观察变量在预测柴油车二氧化碳排放的工作中具有良好表现。Rimpas 等^[27]通过 OBD-II 协议获取车辆的运行数据,监测运行状态,估算瞬时油耗,并通过实验发现驾驶行为对油耗有显著影响。Rosero 等^[28]通过将 PEMS 与 OBD 数据采集系统的数据结合,获得了瞬态发动机数据,绘制了发动机效率-排放图,使用这些数据评估发动机在现实条件下的性能,并创建了发动机的油耗和排放逐秒预测模型。Pavlovic 等^[29]通过对比实验室循环工况、PEMS 与 OBD 数据,验证欧盟法规下 OBD 检测数据精度,发现 OBD 燃油消耗测量数据可以满足欧盟精度要求,但精度受车速、加速度和行程动态性影响大。Rivera-Campoverde 等^[30]利用 OBD 采集多种道路上车辆的运行数据,计算车辆的污染物瞬时排放率,并使用 Savitzky-Golay 平滑算法消除异常值,结合 PEMS 数据,利用机器学习算法,实现了对实际驾驶条件下污染物排放的准确估算和分析。

OBD 设备的普遍安装及其数据集远程传输的实现,降低了数据采集成本,使得车辆大规模运行数据集的获取成为可能。但由于网络延迟、信号质量差等原因,OBD 数据集的质量难以保证,在具体使用中的有效性偏低,且需要与其他

数据集结合，以更全面的评估车辆排放。

1.3 研究目的与内容

本研究旨在通过融合 OBD 数据和道路特征数据，结合机器学习算法，构建驾驶工况-道路环境耦合的时空数据集，综合考虑工况与环境对重型柴油货车 NO_x 排放的影响，获取重型柴油货车真实环境下准确的 NO_x 排放因子，以及基于时空数据集构建 NO_x 瞬态排放预测模型，用于中微观高分辨率 NO_x 排放清单的建立。这种基于近实时数据获取的 NO_x 排放因子以及排放预测模型，突破了基于统计或者实验数据建立的排放清单的滞后性和不准确性，以更加准确的 NO_x 排放数据反映实际排放状况，从而为生态环境部门制定空气质量持续改善方案、重型柴油货车污染防治政策等提供精准数据支持，助力科学、精准防治大气污染。同时，丰富了移动源污染排放核算的研究方法，为后续相关研究在考虑真实环境因素和瞬态排放获取方面提供新思路与参考。本文的具体研究如图1.2所示。

(1) 本文通过融合基于 OBD 设备获取的车辆运行数据和道路信息，构建时空数据集，弥补以往统计或实验室数据集数据量少、获取成本高的不足，并基于此获取重型柴油货车真实环境下的基准 NO_x 排放因子以及道路类型差异化的 NO_x 排放因子，提高 NO_x 排放因子的真实性与准确性，填补了国六重型柴油货车 NO_x 排放因子的空白。其中，时空数据集的构建主要涉及 OBD 数据、工况数据以及道路信息数据的获取与融合。首先，结合线性插值、多层感知机 (Multilayer Perceptron, MLP) 和随机森林算法 (Random Forest, RF)，提出一种分阶段插值策略，对筛选后的 OBD 数据集进行填补，在保留原始数据特征的基础上，提高 OBD 数据的完整性与可用性。接着，利用 MOVES 模型的 VSP 计算方法，获取重型柴油货车的 VSP 工况数据，并基于 VSP 进行工况聚类划分。最后，利用全球行政区划数据集 (Database of Global Administrative Areas, GADM) 的中国省级行政边界数据和开放地图资源 (OpenStreetMap, OSM) 的路网地图，结合车辆运行范围的经纬度边框，获取车辆的道路特征信息。将以上数据集融合，获取驾驶工况-道路环境耦合的时空数据集。基于此数据集，计算获取真实环境下国六重型柴油货车的基准 NO_x 排放因子为 1.64g/km ，并结合道路信息获取快速路、主干路、次支路上的综合排放因子。对比发现，快速路排放因子 (1.82g/km) 较主干路 (1.28g/km) 与次支路 (1.24g/km) 高出 29.7%-31.9%，体现了道路导致的 NO_x 排放差异性。

(2) 本文通过集成轻量级梯度提升机算法 (Light Gradient Boosting Machine, LightGBM) 和极端梯度提升算法 (eXtreme Gradient Boosting, XGBoost)，基于时空数据集，建立了重型柴油货车的 NO_x 瞬态排放预测模型，突破了传统排放模

型无法获取瞬态排放的局限，获取更精准的 NO_x 排放数据，并尝试用于重型柴油货车高分辨率 NO_x 排放清单的建立。基于时空数据集，借助相关系数和 SHAP 值（SHapley Additive exPlanations）分析车辆多维特征与 NO_x 排放的关系，通过特征消融，获取与 NO_x 瞬态排放最为相关的八个车辆特征。考虑到多维时空特征间存在非对称交互与高阶非线性耦合特性，本文综合利用 LightGBM 在大规模数据集上的高效性以及 XGBoost 的 L1/L2 正则化约束抑制过拟合，并通过堆叠泛化（Stacking）进行集成，最终构建了重型柴油货车的 NO_x 瞬态排放预测模型，即 SLXGB- NO_x 排放模型。该模型相较于基础模型，在预测准确性提升了约 13%。另外，通过该模型建立了成都市重型柴油货车的高分辨率 NO_x 排放清单，分析了其 NO_x 排在时空上的分布特征，即时间维度呈现显著的“双峰”特征，日间排放贡献率高达 79.9%，空间上在绕城高速聚集，呈现环状分布，与已有研究中成都市 NO_x 排放特征比分布相似，验证了排放清单的可靠性。

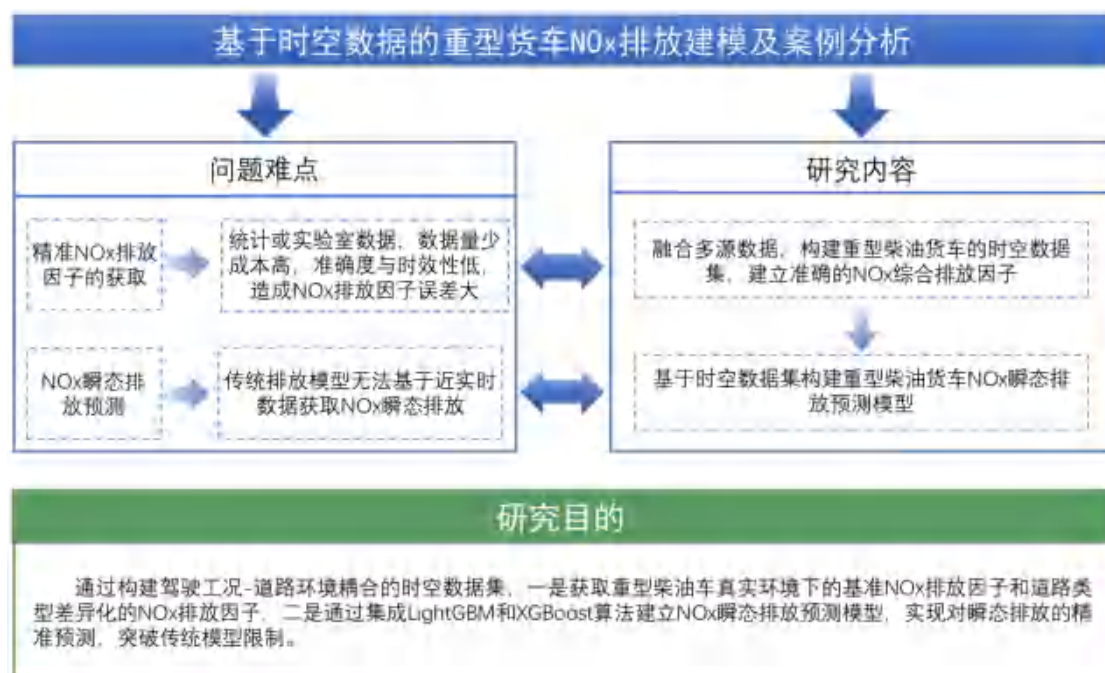


图 1.2 本文主要研究内容

1.4 本文组织结构

本文使用五个章节的内容对这次的研究进行了详细介绍。

第一章 绪论。该章节首先介绍了本文的研究背景和意义，然后围绕传统排放模型和基于 OBD 数据的排放研究两个方面，开展了国内外研究现状的介绍与分析，指出了基于多源数据进行排放研究这一趋势，最后详细阐述了本文研究的目的和内容，以及各个章节的安排。

第二章 相关理论基础。该章节首先讲述了回归算法的概念原理，其次介绍

了用于实现插值和建立 NO_x 排放模型的基于树的集成学习算法以及组合方法的原理与优势,最后介绍了用于数据检验与评价相关方法的原理与计算公式,为后续构建时空数据集、建立重型柴油货车 NO_x 排放模型和模型效果评价提供理论基础。

第三章 基于时空数据的真实 NO_x 排放因子。该章节首先对 OBD 数据进行采集与处理,并在插值阶段提出了一种分阶段数据插补方法,对具有一定连续性的数据使用线性插值,对 SCR 系统的 NO_x 传感器输出值利用多层感知机与随机森林算法构建插值模型。其次,基于 OBD 数据,利用 MOVES 模型的 VSP 计算方法,获取工况数据。然后,结合 GADM 和 OSM 资源获取路网地图,根据 GPS 车辆轨迹数据,获取车辆的道路信息。通过时间与车辆 VIN 码的匹配,将数据集融合,获取驾驶工况-道路特征耦合的时空数据集,计算获取基于真实环境数据的 NO_x 排放因子以及不同道路下的综合排放因子。

第四章 基于堆叠泛化的 NO_x 瞬态排放预测方法。该章节基于第三章构建的重型柴油货车时空数据集,首先利用相关系数分析,选取与 NO_x 瞬时排放相关的参数。其次,基于 XGBoost 算法,构建 NO_x 基础排放模型,并借助 SHAP 值分析参数的重要性,并通过单变量特征消融,获取八个与 NO_x 瞬态排放最为相关的特征。之后,利用堆叠泛化集成 LightGBM 和 XGBoost 算法,构建 SLXGB- NO_x 瞬态排放模型。通过与其他基准模型对比,SLXGB- NO_x 模型在预测准确性与时间效率上都表现更佳。最后,利用 SLXGB- NO_x 瞬态排放预测模型,结合成都市重型柴油货车远程监控平台的 OBD 数据,获取成都市重型柴油货车的高分辨率 NO_x 排放清单,分析获取该市重型柴油货车的 NO_x 排放在时空上的分布特征。

第五章 结论与展望。该章节对本次研究的主要工作进行了总结,分析了其中存着的不足之处,并提出了未来可以进一步探索与完善的方向。

第2章 相关理论基础

在本章节中，对构建时空数据集和 NO_x 排放模型过程中所用方法的基本概念和原理进行介绍。首先是对岭回归、多层感知机两个回归算法的介绍。其次是对随机森林、极端梯度提升、轻量级梯度提升以及堆叠泛化这些基于树的集成模型以及组合方法的讲述。另外，还介绍了用于数据检验与评价的方法，包括判断数据分布是否一致的 K-S 检验、用于评估参数重要性的递归特征消除与 SHAP 值，以及相关系数 R^2 和均方根误差 RMSE 这两个评价指标的计算原理，为后续的工作打好理论基础。

2.1 回归算法基础

2.1.1 岭回归

在传统线性回归中，当特征间存在高度多重共线性时，参数估计易受干扰，导致模型过拟合，泛化能力下降。岭回归针对这一缺陷改进，在损失函数中引入 L2 正则化项，通过约束参数控制模型复杂度。该算法既适用于直接回归预测，也常作为元学习器融入集成模型，提升整体稳健性，广泛应用于需要平衡拟合与泛化的场景。在该算法的目标函数表达式 (2.1) 中，正则化参数 λ 是关键调节因子。当 λ 较小时，模型侧重数据拟合； λ 较大时，压缩参数简化模型，避免过拟合。

$$\mathbf{J}(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \theta^T x_i)^2 + \lambda \|\theta\|_2^2. \quad (2.1)$$

2.1.2 多层感知机

多层感知机 (Multilayer Perceptron, MLP)^[31] 是一种前馈人工神经网络，其设计灵感源自对大脑神经元信息处理机制的仿生。MLP 通过模拟神经元的连接与信号处理过程，构建起数据特征与目标结果之间的映射关系。凭借强大的非线性拟合能力，MLP 在数值预测、模式识别等领域展现出独特优势，能够通过优化网络权重与偏置参数，深度挖掘数据中隐含的复杂规律，为复杂问题提供精准的分析与预测支持。

MLP 通常由一个输入层、一个或者多个隐藏层以及一个输出层构成，信息只能单向的从输入层到输出层传播，结构如图2.1所示。输入层承担数据接入功能，将数据样本转化为特征向量输入模型。作为特征提取的核心模块，隐藏层通过权重 $W^{(l)}$ 、偏置 $b^{(l)}$ 对输入信号进行线性变换，再经过激活函数（如 ReLU、Sigmoid）对线性变换结果进行非线性映射，提取复杂特征。在输出层，根据任

务类型（分类或回归）设置神经元数量。若为回归任务，输出层神经元直接输出连续值；若为分类任务，则通过激活函数将输出转化为概率分布。每个输出神经元整合前层信号，经加权求和与激活函数处理后生成最终输出值。

以 σ 为激活函数， $h^{(l)}$ 代表第 l 层的激活值， $W^{(l)}$ 则为第 l 层的权重矩阵， $b^{(l)}$ 是第 l 层的偏置向量，MLP 的数学表达式如公式 (2.2) 所示。该公式递归描述了 MLP 从输入层到输出层的信号传递过程，体现了网络通过多层非线性变换逼近复杂函数的本质特性。通过反向传播算法优化 $W^{(l)}$ 和 $b^{(l)}$ ，MLP 能够最小化预测值与真实值的误差，最终实现对复杂数据模式的精准建模。

$$h^{(l)} = \sigma(W^{(l)}h^{(l-1)} + b^{(l)}). \quad (2.2)$$

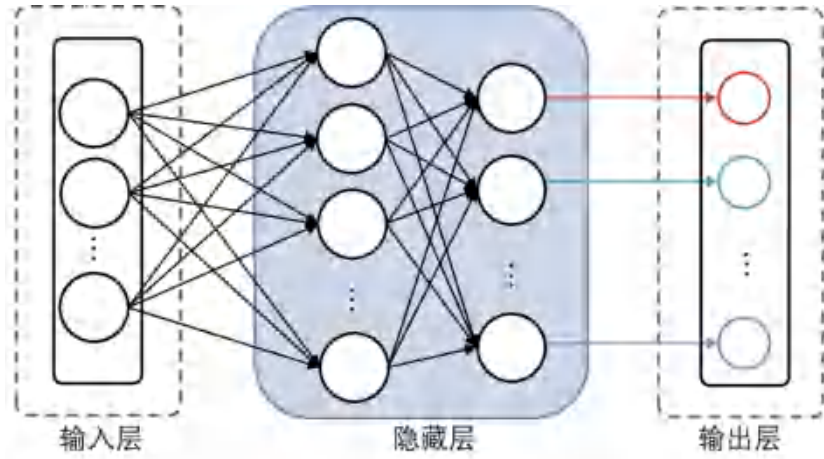


图 2.1 多层感知机结构图

2.2 基于树的集成模型及组合方法

2.2.1 随机森林算法

随机森林算法 (Random Forest, RF)^[32-33] 是一种基于决策树的集成学习算法，兼具分类与回归任务的处理能力。其核心机制是通过构建多棵决策树形成“森林”，将单棵决策树的预测结果通过投票（分类任务）或均值回归（回归任务）进行集成，输出最终的预测结果，实现对数据的精准分析。此外，每棵决策树独立训练且随机选取子样本，从机制上降低过拟合风险。相较于传统回归方法，该算法能有效捕捉数据间复杂的非线性关系，凭借高鲁棒性与泛化能力，在数据预测中展现突出优势。

在构建过程中，随机森林有两大关键的随机步骤体现了其随机性：一是采用有放回随机抽样方法获取训练样本，使每棵决策树基于不同样本集训练，增强模型多样性；二是在决策树节点分裂时，从随机选取的部分特征中寻找最优分裂方式，进一步强化树与树之间的差异性。这个双重随机性设计带来的随机性使得每

棵决策树在训练过程中形成独特的特征识别模式，最终通过集成策略减少模型方差，降低过拟合风险。随机森林算法的典型实现流程如图2.2所示。

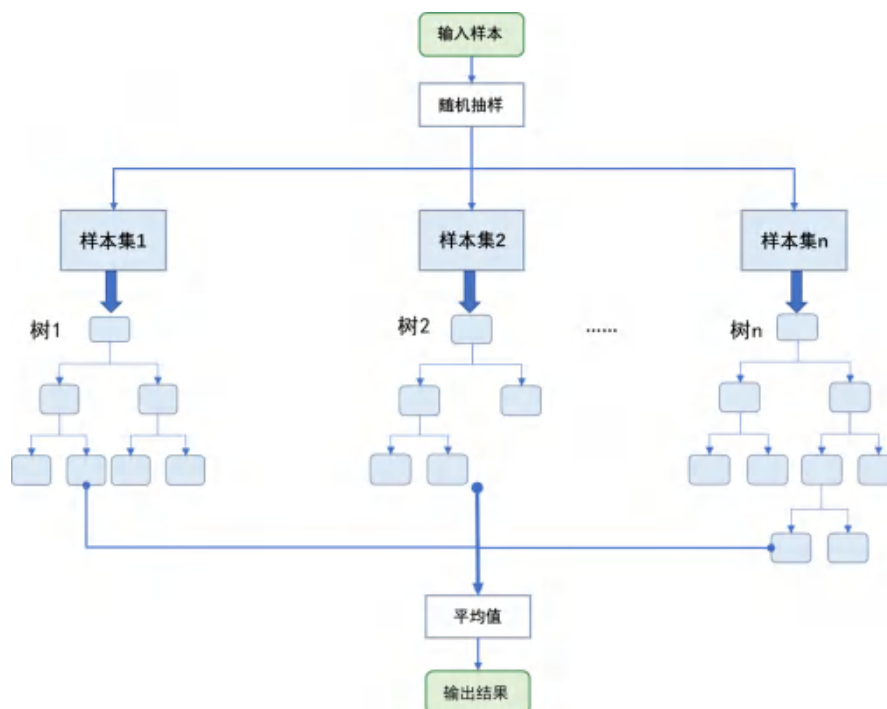


图 2.2 随机森林算法过程图

2.2.2 极端梯度提升算法

极端梯度提升算法（eXtreme Gradient Boosting, XGBoost）由 Chen 等^[34]于 2016 年提出，是基于梯度提升决策树（Gradient Boosting Decision Tree, GBDT）框架的优化算法模型，其核心思想是通过迭代地训练多个弱学习器（通常是决策树），逐步减少模型的预测误差。在每次迭代中，利用损失函数的负梯度来拟合新的决策树，并通过正则化技术（如限制树的深度、对叶子节点的权重进行 L1 和 L2 正则化）来控制模型的复杂度，从而有效防止过拟合。

相较于传统的 GBDT，XGBoost 的核心改进体现在三个方面：一是在计算损失函数时，利用二阶展开的泰勒公式，更精准地对损失函数进行近似；二是在模型中加入正则化参数，来控制模型的复杂性，抑制过拟合问题的产生；三是支持基于特征维度的并行运算，提升模型训练效率。

XGBoost 的构建流程包含三个核心阶段：

（1）树结构的生成：采用贪心算法逐层优化树结构，通过增益（Gain）函数选择特征分裂点。以公式 (2.3) 的 Gain 函数为准则，计算分裂前后损失函数的差值，并通过最大化这个差值完成特征划分。同时，为防止模型过拟合，还会通过限制最大深度（max_depth）这一参数，对树的生长深度进行约束。一旦树的深

度达到预设值，便停止生长，以此控制模型复杂度。

$$\text{Gain} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma, \quad (2.3)$$

其中， G_L/G_R 、 H_L/H_R 分别为左右子节点的梯度和二阶导数之和， λ 为正则化参数，用于控制模型复杂度， γ 为分裂复杂度。

(2) 目标函数优化：这里的目标函数通过融合损失函数与正则化项来实现优化，具体形式如公式 (2.4)。从公式构成来看，其中的二阶泰勒近似部分反映了当前模型预测值与真实值之间的差异情况，通过对这部分的优化，模型能够不断逼近真实结果；结构正则化项则是用于对模型的结构进行约束，防止模型出现过拟合。

$$\text{Obj}^{(t)} = \underbrace{\sum_{i=1}^n \left[g_i f_i(\mathbf{x}_i) + \frac{1}{2} h_i f_i^2(\mathbf{x}_i) \right]}_{\text{二阶泰勒近似}} + \underbrace{\gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2}_{\text{结构正则化项}}, \quad (2.4)$$

$$\begin{cases} g_i = \frac{\partial}{\partial \hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}), & \text{一阶梯度,} \\ h_i = \frac{\partial^2}{\partial (\hat{y}_i^{(t-1)})^2} l(y_i, \hat{y}_i^{(t-1)}), & \text{二阶梯度,} \end{cases} \quad (2.5)$$

其中， t 表示当前迭代的轮次， n 为样本量， T 为树的叶子节点数量。 w_j 为第 j 个叶子的权重值， γ 、 λ 是正则化系数。 g_i 代表关于 $\hat{y}_i^{(t-1)}$ 对损失函数 $l(y_i, \hat{y}_i^{(t-1)})$ 求一阶导数所得的一阶梯度，体现了损失函数在当前点处的变化趋势，指引模型优化方向； h_i 代表关于 $(\hat{y}_i^{(t-1)})^2$ 对损失函数 $l(y_i, \hat{y}_i^{(t-1)})$ 求二阶导数所得的二阶梯度，它反映了损失函数变化的曲率信息，帮助模型更精准地调整拟合程度。

(3) 特征重要性评估：特征重要性主要通过加权增益 (Gain)、覆盖率 (Cover)、频率 (Frequency) 三个方面解释。其中，Gain 是主评估指标，体现特征在所有树中的平均增益贡献，Cover 则表现了特征被用为分裂结点的样本覆盖率，Frequency 则是特征在分裂时被使用的相对次数。通过这三个维度的综合评估，能够全面了解每个特征在模型中的重要性，为特征筛选和模型优化提供依据。

XGBoost 的优势在于其出色的预测精度和广泛的适用性，能够在复杂非线性建模任务中展现出显著优势，并且支持并行计算和分布式训练，能够高效处理大规模数据集。它在许多机器学习竞赛和实际应用中表现优异，尤其是在结构化数据的建模任务中。

XGBoost 也存在一些局限性。首先，模型调参相对复杂，需要一定的经验和技巧才能找到最优的超参数组合。其次，XGBoost 对异常值和噪声较为敏感，因此，在数据预处理阶段需要进行适当的清洗和标准化处理。此外，XGBoost 尽管支持并行计算，但在处理极高维数据时，训练时间仍然可能较长。

2.2.3 轻量级梯度提升机算法

轻量级梯度提升机算法 (Light Gradient Boosting Machine, LightGBM)^[35] 作为一种基于决策树的分布式并行梯度提升框架, 其核心优势体现在高效训练速度、低内存消耗和高精度预测性能。其目标函数的表达式如公式 (2.6) 所示。

$$\text{Obj}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t), \quad (2.6)$$

其中, $l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$ 是损失函数, 用于衡量真实值 y_i 与前 $t-1$ 次迭代预测值 $\hat{y}_i^{(t-1)}$ 加上当前迭代函数 $f_t(x_i)$ 之间的差异。另外, 正则化项为 $\Omega(f_t) = \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2$, 通过 L1 与 L2 正则化约束模型的复杂度, 防止过拟合问题的产生。

相较于 XGBoost 算法, LightGBM 通过改进样本采样算法和决策树构建机制, 实现了在不降低准确率的同时大幅度提升模型的训练效率, 其核心依托于以下三个关键技术:

(1) 单边梯度采样算法 (Gradient-based One-Side Sampling, GOSS): GOSS 是一种基于梯度统计特性的数据优化方法, 其核心机制在于优先保留对模型训练具有显著价值的数据实例。根据梯度与信息增益的正相关性原理, 梯度幅值较大的样本通常在决策树构建过程中产生更大的信息增益量。因此, GOSS 采用非对称处理策略: 在保持大梯度样本完整性的同时, 策略性地减少对小梯度数据的计算关注。这种选择性采样机制既保留了关键训练信息, 又有效降低了计算复杂度, 尤其适用于海量数据场景下的模型优化。

(2) 直方图算法 (Histogram algorithm): 在构建决策树时, 该算法的基本原理是先将连续的特征转化为 k 个箱, 每个箱对应一个整数值。同时包含 k 个箱的直方图。在遍历数据的过程中, 会把离散化后的整数值作为索引, 于直方图中对相关统计量进行累加。待遍历结束, 直方图即获得所需统计量, 后续只需依据离散化后的整数值遍历即可找到最优分割点, 无需再遍历原始特征值。

(3) 带有深度限制的逐叶生长策略: 该策略采用双重控制机制, 在决策树构建过程中动态选择分裂增益最大的叶节点进行扩展, 并设置深度阈值约束树结构生长, 防止叶子过多引发的过拟合问题。

相较于传统 GBDT 使用的逐层生长策略, 本方法通过增益导向的节点优选机制, 可以提升模型的精度, 降低误差, 同时对分裂增益较低的叶节点不进行处理, 降低运算的开销, 进一步提升算法的效率。其中, GBDT 中树的逐层生长策略与 LightGBM 的逐叶生长策略分别如图2.3的 (a)、(b) 所示, 其中逐叶生长模式呈现出显著的非对称分支特征。

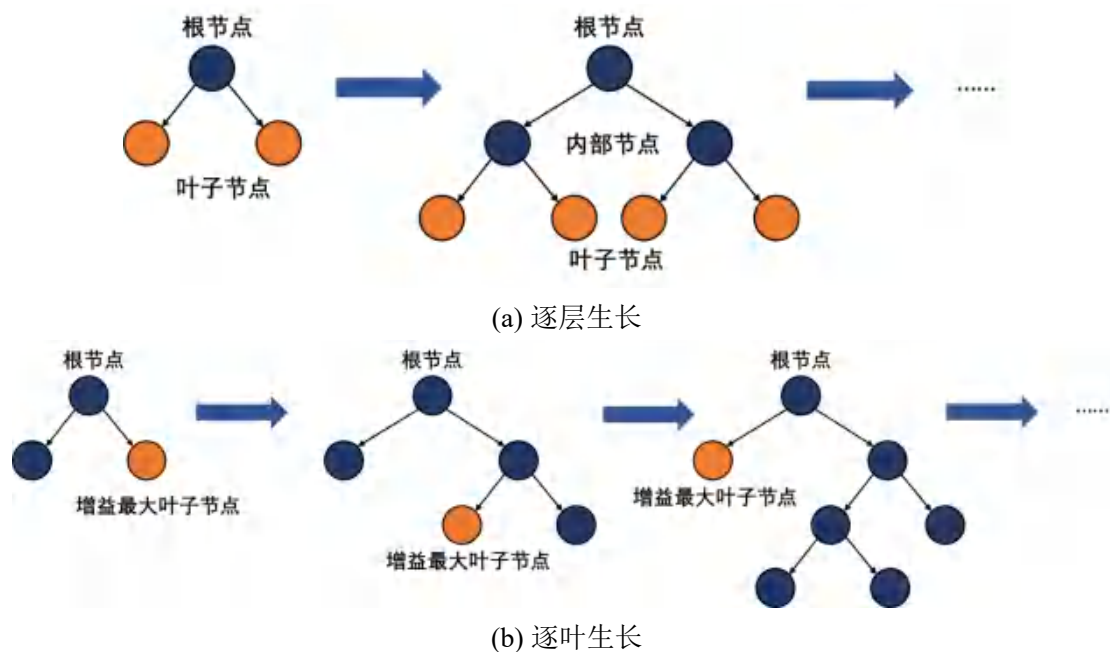


图 2.3 树的生长策略对比

2.2.4 堆叠泛化

堆叠泛化 (Stacking) 是一种元学习方法，通常采用两层架构设计。其中，第一层是使用多个不同的模型作为基学习器，这些基学习器构建了基模型层；第二层则是训练一个元模型，专门用于组合基学习器的输出结果，最终完成预测任务。

作为一种集成学习方法，Stacking 的优缺点也较为明显。其优点是能够充分利用不同模型的优势和特点，通过融合多个基础模型的预测结果，显著提高整体模型的预测精准度和泛化能力。同时由于可以集成不同类型的模型，Stacking 可以非常灵活的适应不同类型的数据和问题。

Stacking 存在的缺陷是由于需要依次训练多个基模型，再基于基模型的输出训练元模型，因而需要更多的资源和时间，同时增加了模型调参的复杂度。另外，当基础模型过于复杂或者训练集过小时，基模型容易出现过拟合，且这种过拟合问题会通过输出传递给元模型，最终导致整个 Stacking 模型的泛化性能下降。图2.4展示了 Stacking 模型架构的一般结构示意图。

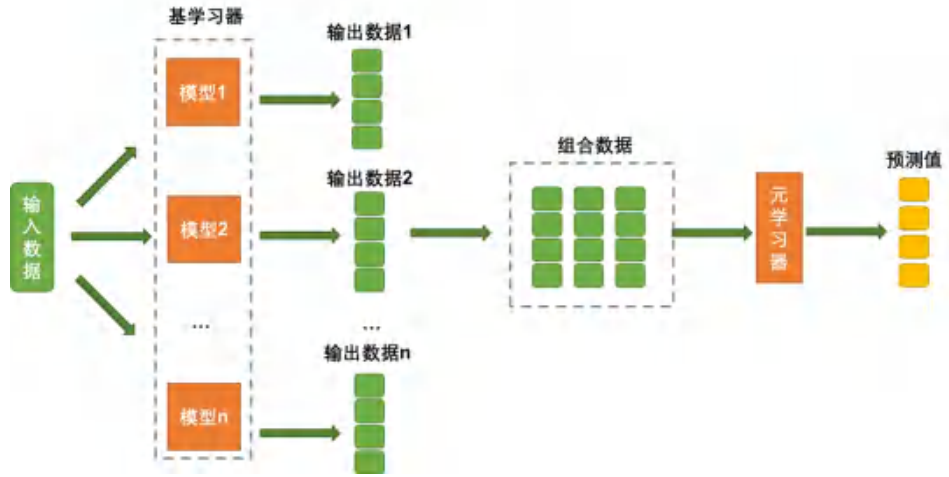


图 2.4 Stacking 结构示意图

2.3 数据检验与评价指标

2.3.1 K-S 检验

K-S 检验是一种非参数假设检验方法，无需对数据分布形式（如正态性、方差等）做严格假定，主要有两种应用场景：一是判断一组样本数据是否符合某一个已知理论分布，如正态分布、均匀分布；二是用于检验两组独立的样本是否具有相同的分布，判断其分布差异是否显著。

对于第一种单样本数据检验，其原理是：将样本升序排列后，计算每个数据点的累计频率，获取样本的经验累计分布函数（Empirical Cumulative Distribution Function, ECDF）。同时，根据假设的理论分布，计算理论累计分布函数（Cumulative Distribution Function, CDF）。然后，比较每个数据点的 ECDF 和 CDF，获取二者差值绝对值的最大值 D ，再对比 D 与 K-S 检验临界表。若 D 大于临界值，则拒绝原假设，认为样本不服从假设的理论分布；反之，则不能拒绝原假设。 D 值计算公式如 (2.7)，其中的 $F_n(x)$ 为经验分布函数， $F(x)$ 为理论分布函数。

$$D = \max_x | F_n(x) - F(x) |. \quad (2.7)$$

对于第二种双样本数据检验，原理为：分别建立两组样本的经验分布，计算各自的 ECDF。然后，合并两组数据并排序后，计算两组 ECDF 的差值，同样取最大的绝对值 D ，与临界值对比或者计算 p 值得出结论。若 D 超过临界值且 p 值小于显著性水平 α ，则拒绝“两组数据分布相同”的原假设；反之，不能拒绝。假设 $X = \{x_1, x_2, \dots, x_n\}$ 和 $Y = \{y_1, y_2, \dots, y_m\}$ 为两组样本数据，合并两组数据并排序，得到有序序列 $z_1 \leq z_2 \leq \dots \leq z_{n+m}$ ，进而在每一点 z_i 处计算每组数据的 ECDF，分别为 $F_n(z_i)$ 和 $G_m(z_i)$ ，则 D 的公式如下：

$$D = \max_{1 \leq i \leq n+m} | F_n(z_i) - G_m(z_i) |. \quad (2.8)$$

上述累计分布函数的一般表达式为公式 (2.9)。其中, n 为样本量, X_i 为第 i 个样本的值, $\mathbf{I}(\cdot)$ 为指示函数, 满足 $X_i \leq x$ 的条件则取 1; 反之, 取 0。

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(X_i \leq x). \quad (2.9)$$

2.3.2 递归特征消除

递归特征消除 (Recursive Feature Elimination, RFE) 是一种基于模型的特征选择方法, 核心在于从原始特征集中挑选出最优特征子集。RFE 通过迭代训练模型, 根据模型输出的特征重要性指标, 每次剔除重要性得分最低的特征, 循环此过程直至确定最优特征组合。

RFE 的优势在于: 可以与多种算法灵活组合, 适配不同数据特性与任务需求; 能够去除冗余和噪声特征, 降低特征维度, 有效减少过拟合风险, 提升模型泛化能力; 通过筛选出关键特征, 使得模型内部机制更易理解。但其局限性在于: 特征重要性评估明显依赖于基模型, 且需要反复训练模型以迭代提出特征, 消耗资源较多。

2.3.3 SHAP 值

SHAP 值 (SHapley Additive exPlanations) 是一种源自合作博弈论 Shapley 值理论的模型可解释性技术, 其核心功能在于评估每个特征对特定样本预测结果的影响程度。通过这种量化分析方法, 能够清晰展现模型决策过程中不同特征的贡献度分布, 从而提供清晰的模型决策依据。SHAP 值的核心思想源自 Shapley 值的公平分配原则, 满足可加性 (Additivity)、对称性 (Symmetry) 和零贡献性 (Null Player) 公理。在机器学习场景下, 每个输入的特征可以看作博弈中的“玩家”, 模型预测结果与基线预测的差值则类比为“总收益”。SHAP 值通过计算每个特征在所有可能组合中的边际贡献, 公平分配其对预测结果的影响。

对于一个机器学习模型的某个特征 i 的预测结果, SHAP 值计算会考虑所有可能的特征组合, 拆解为各个特征对该结果的贡献之和。其基本计算公式如下:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} \left[\mathbf{E}[f \mid X_{S \cup i} = x_{S \cup i}] - \mathbf{E}[f \mid X_S = x_S] \right], \quad (2.10)$$

其中, N 是所有特征的集合 ($n = |N|$ 为特征总数), S 是一个不包含特征 i 的特征子集, f 为机器学习模型, 输入为特征向量 X , 即 $f(X)$ 简写为 f 。 $\mathbf{E}[f \mid X_S = x_S]$ 表示已知特征子集 S 取值为 x_S 时, 模型输出的条件期望。权重项 $\frac{|S|!(n - |S| - 1)!}{n!}$ 表示在所有可能的特征加入顺序中, 子集 S 出现在特征 i 之前的概率, 用于对特征 i 的边际贡献 $\mathbf{E}[f \mid X_{S \cup i} = x_{S \cup i}] - \mathbf{E}[f \mid X_S = x_S]$ 进行加权平均。

SHAP 值会为每个特征生成一个数值来表示其对单个预测结果的影响，其绝对值越大，表示该特征对预测结果的影响越显著。当 SHAP 值为正值时，说明该特征对预测结果的贡献是正向的，可以促使模型预测更高的输出；反之，则对预测结果的贡献是负向的，会使模型预测更低的输出。另外，由于遍历所有子集的计算复杂度是指数级别，在高维数据上直接计算不现实。因此，SHAP 提供了多种近似算法，如基于树模型的快速计算方法、通过随机采样不同特征子集来减少计算量的方法等，适用于大规模数据集和多种复杂模型。

2.3.4 相关系数

相关系数 R^2 (Coefficient of determination)，也称为决定系数，用于衡量一元多项式回归方程拟合度的高低，或者说表示回归方程估测的可靠程度。在回归分析中，目标变量观测值相对于其均值的总离差 (Total Sum of Squares, TSS) 可划分为两个关键组成部分：一部分是回归平方和 (Explained Sum of Squares, ESS)，反映了模型对数据变异的解释能力；另一部分为残差平方和 (Residual Sum of Squares, RSS)，表征了模型未能捕捉到的随机误差部分。具体计算公式如下：

$$\begin{cases} ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \\ RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2, \\ TSS = ESS + RSS, \end{cases} \quad (2.11)$$

其中， $\hat{y}_i$ 是模型的预测值， y_i 是观测值， $\bar{y}$ 是观测值的均值。

R^2 的数学表达式如公式 (2.12) 所示，该指标取值范围介于 0 到 1 之间。当 R^2 值趋近于 1 时，表明实际观测数据点与样本回归直线的拟合程度越高，反映出模型对数据的拟合优度越强，即模型对因变量变异的解释比例越高；反之，当 R^2 值趋近于 0 时，则说明模型对数据的拟合效果较差，其所能解释的因变量变异比例也相应较低。这种量化评估方式为模型拟合效果的优劣提供了直观的判别依据。

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}. \quad (2.12)$$

2.3.5 均方根误差

均方根误差 (Root Mean Squared Error, RMSE) 是衡量回归模型预测值与真实值之间误差的一种常用指标。其计算公式为：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2.13)$$

其中， n 是样本数量， y_i 是第 i 个样本的真实值， $\hat{y}_i$ 是第 i 个样本的模型预测值。

RMSE 作为残差平方和的平均值平方根，其计算过程包含对预测误差的平方化处理，这一特性使得该指标对较大误差赋予更高权重。该指标的数值大小直接反映模型预测值与真实值的偏离程度：当 RMSE 值较小时，表明模型预测结果与实际观测值的一致性较好，体现出更高的预测精度；反之，较大的 RMSE 值则意味着模型预测结果与真实值存在较大偏差，预测性能有待提升。

2.4 本章小结

本章对本研究工作过程中会使用到的方法进行了详细讲述。首先是按照方法类别，介绍了常用于回归预测任务的岭回归以及多层感知机方法。其次，为了处理后续数据之间复杂的耦合关系，这里选择了随机森林、极端梯度提升、轻量级梯度提升机以及堆叠泛化方法，并对它们的原理以及优缺点进行介绍。最后，介绍了用于数据检验以及参数分析检验的方法，包括递归特征消除、K-S 检验、SHAP 值、决定系数以及均方根误差。为后续的一系列工作打好理论基础。

第3章 基于时空数据的真实 NO_x 排放因子

本章构建了重型柴油货车驾驶工况-道路特征耦合的时空数据集，获取实际道路上的 NO_x 国六排放因子，分析道路特征等多方面因素对排放的影响。研究通过车载诊断系统采集车辆运行数据，结合线性插值、多层感知机以及随机森林算法完成运行数据的分阶段插补；基于 MOVES 模型的 VSP 计算方法，获取工况数据；根据 GADM 和 OSM 资源获取路网地图，获取道路信息。通过以上多源数据集融合形成时空数据集。在此基础上，获取了重型柴油货车实际道路上的基准 NO_x 排放因子以及不同道路类型的综合 NO_x 排放因子，填补了真实环境下的 NO_x 国六排放因子的缺失，且对比发现了道路特征对排放因子造成的差异化。

3.1 OBD 数据采集与处理

3.1.1 OBD 数据采集

OBD 数据采集装置主要包括数据采集单元、无线传输系统以及数据存储单元。在车辆运行时，数据采集单元可以获取车辆发动机的相关运行参数，包括车速、发动机燃料流量、发动机转速等。对于配备了 SCR 尾气后处理系统的车辆，OBD 还会采集 SCR 系统的工作状态，包括 SCR 系统出入口温度、上下游 NO_x 浓度等，以此监测 SCR 处理系统的运行状况以及车辆的尾气排放，确保能有效降低尾气中 NO_x 的排放。之后，将数据存储单元中的数据包通过无线传输系统发送到中心服务器，进行存储与解析，计算机终端用户便可通过访问数据服务器获取车辆的运行数据。

OBD 数据采集终端的结构以及数据传输过程示意图如图3.1、3.2所示。在车辆排放研究中，常用的 OBD 数据字段包括车速、发动机数据、SCR 传感器数据、经纬度等。表3.1则展示了本次研究选用的 OBD 数据集的具体参数字段名称以及单位。

3.1.2 数据预处理

由于 OBD 数据按照采集时间存放，会有多辆车数据混合在一个数据表或者一辆车的数据跨文件保存的情况。因此，为了保证 OBD 数据的质量，对所获取的全部数据汇总，并进行初步的数据筛选与清洗操作。本次研究的对象是配备有 SCR 类型尾气处理装置、质量在 12t 以上的国六重型柴油货车，根据重型货车远程监控技术规范^[36]对 OBD 数据的要求以及多次的数据处理实验，本次重型柴油货车的 OBD 数据筛选与清洗流程如图3.3所示。具体的数据筛选与清洗标准如



图 3.1 OBD 数据采集终端结构



图 3.2 OBD 数据传输过程

下：

（1）车辆准入条件：类型为 N3、燃料类型为柴油、排放标准为国六、总质量在 12t 以上、配备了 SCR 尾气处理装置的重型载货汽车。

（2）时间合理性检验：根据 OBD 数据先采集存储，然后发送至数据服务器的特点，对于数据集中采集时间晚于接收时间的无效数据要进行舍弃。

（3）数据格式转换：OBD 数据以十六进制的形式传输到数据服务器中，需要对数据进行解码，然后利用精度和偏移量实现数据校准，获取真实值，一般公式如下：

$$\text{Value} = \left(\sum_{i=1}^n \text{hex2dec}(D_i) \times 2^{8(n-i)} \right) \times \text{Accuracy} + \text{Offset}, \quad (3.1)$$

其中，Value 为转换后的真实数据值， D_i 代表第 i 字节的十六进制值， $\text{hex2dec}(\cdot)$ 实

表 3.1 OBD 数据集字段说明

名称	字段名	单位
车速	speed	km/h
大气压力	atm	kPa
发动机净输出扭矩	neot	%
摩擦扭矩	eft	%
发动机转速	erotation	rpm
发动机燃料流量	efflow	L/h
SCR 上游 NO _x 传感器输出值	scrnoxUP	ppm
SCR 下游 NO _x 传感器输出值	scrnoxDn	ppm
反应剂余量	reactant	%
SCR 入口温度	scrTin	°C
SCR 出口温度	scrTout	°C
DPF 压差	DPF	kPa
油箱液位	tankL	%
进气量	airin	kg/h
发动机冷却液温度	ecolT	°C
累积里程	mile	km
定位状态	locstatus	
经度	lng	°
纬度	lat	°

现十六进制向十进制的转换， n 代表数据占用的字节数，Accuracy 为精度，Offset 为偏移量。

（4）数据质量标准：单车数据量的有效数据帧 ≥ 200 ，且时间跨度 $\geq 300s$ ，关键参数的无效值（包括 NULL、0xFFFF 等）占比 $\leq 20\%$ ，异常值（即超过参数值限定范围）占比 $\leq 20\%$ ，其中关键参数如表3.1所示。

（5）无效值与异常值处理：根据技术规范对 OBD 数据的要求，识别其中的无效值数据，将无效记录标记为缺失值（NaN）。

（6）异常值检测：基于给出的参数合理范围，并进一步利用 IQR 异常检测方法中的第 1 四分位数（Q1）和第 3 四分位数（Q3），进行数据异常筛查，将异常值标记为 NaN。

（7）经纬度修正：基于车辆运行位置变化的连续性，对经纬度差值大、发生位置突变的数据点，结合前后数据的有效经纬度数据进行修正。

原始数据集包含 1300 辆车，包括 N2（3.5t-12t）、N3（>12t）等多种类型，

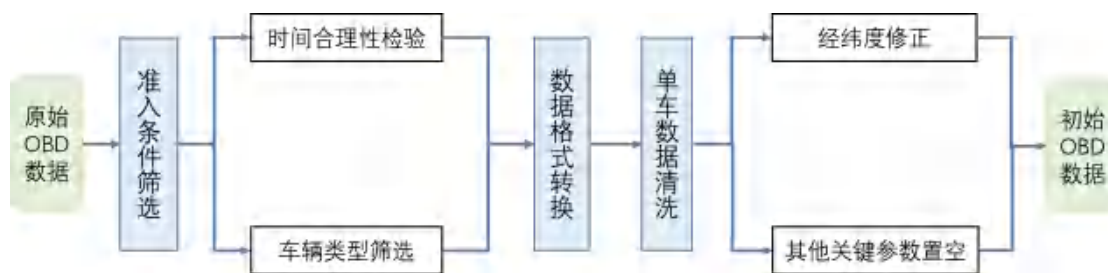


图 3.3 OBD 数据筛选与清洗流程图

经过多层筛选与清洗操作，最终获得包含 973 辆重型柴油货车的可用 OBD 数据集，约占原数据集的 75%。将此数据集中的重型柴油货车，按照车辆总质量所属吨位范围进行划分，并使用数字 1 到 5 对其编号，共分为五种类型的车，每种类型的可用数据量统计结果如表 3.2 所示。

表 3.2 可用数据集统计

车辆类型 (单位)	1(12-16t)	2(16-22t)	3(22-28t)	4(28-40t)	5(40t 以上)	总计
车辆总数 (辆)	3	203	352	411	4	973
数据集 (行)	9,367	503,436	989,821	912,335	9,143	2,424,102

3.1.3 分阶段数据插补

为了保障后续计算与数据分析的准确性，对原始数据及数据清洗过程中产生的缺失值，进行一系列的插值处理，完成数据填补，以确保数据集的连续性与完整性。各关键参数的数据缺失占比如图 3.4，使用主次坐标纵轴区分运行参数与 SCR 传感器的缺失率。其中，SCR 上下游 NO_x 浓度传感器输出值 (scrno_xUP/Dn) 缺失率显著高于其他参数，这主要归因于其容易受到尾气温度、车辆启停状态等工况的耦合影响。基于数据缺失程度与特征差异，实施分阶段插补策略：一是对缺失率 <2% 且具备一定连续性的运行参数进行线性插补，二是对缺失率高且受外部工况约束的 SCR 系统中的 NO_x 浓度传感器数据，利用特征驱动的动态插补算法进行填补。

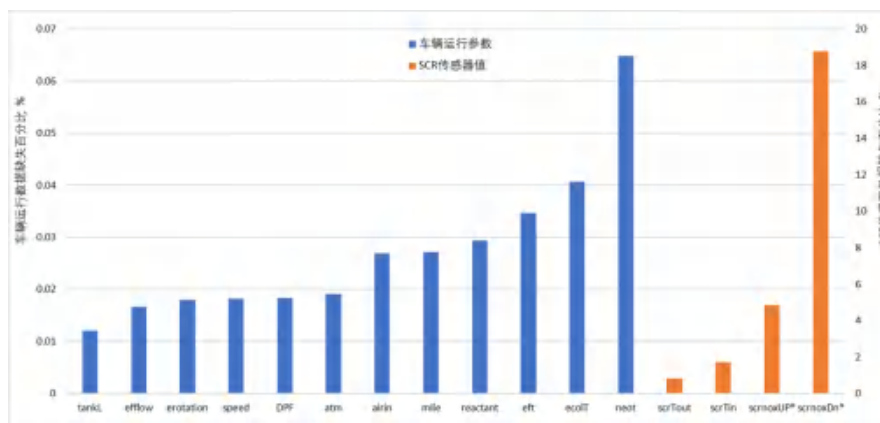


图 3.4 各特征缺失占比

1. 连续性参数插补

(1) 插补方法选择：针对车速、气压等 14 项连续型参数（缺失率 <2%），采用线性插值进行插补效果对比。由于 OBD 数据具有车辆个体差异性，首先按车辆识别码（VIN）分组并按时间戳排序，构建独立数据集。

(2) 插补效果验证方法：为模拟实际缺失场景，对每组数据随机置空 20%（模拟异常占比上限）。通过 K-S 检验插补前后的数据分布差异，并利用决定系数 R^2 量化插补值与真实值之间的相关性。

(3) 实验结果分析：在 14 个连续性参数的插补中，线性插值表现出最优性能。在精度上，线性插值的 R^2 均在 0.95 及以上。且在 K-S 检验上，线性插值后的数据与原始分布一致，如图 3.5 所示。最重要的是，在计算效率上，线性插值的时间复杂度仅为 $O(n)$ ，适合大规模数据的插补。

因此，基于对效率与精度的综合考虑，最终使用线性插值对 14 项参数进行填补，同时保留了车辆原始运行数据的分布特征。填补后的 K-S 检验如图 3.6 所示。

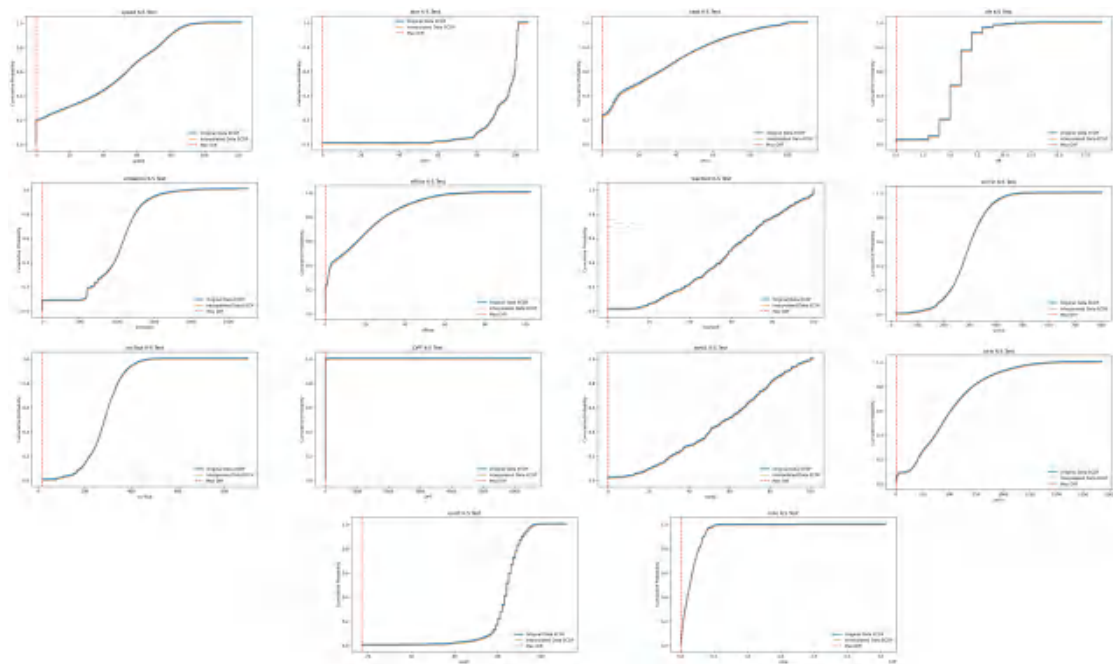


图 3.5 线性插值测试 K-S 检验

2. 基于 MLP-RF 的 NO_x 传感器数据插补

由于 SCR 上下游 NO_x 传感器输出值缺失率高，且使用线性插值会导致其分布发生偏移，数据填补效果不佳。基于 SCR 系统的工作原理及其影响因素，选择与其他特征关联，为 NO_x 传感器值的缺失值建立多变量插值方法。

(1) 多维度相关性分析

首先，基于对重型柴油货车 NO_x 排放的产生原理及其影响因素的深入剖析，

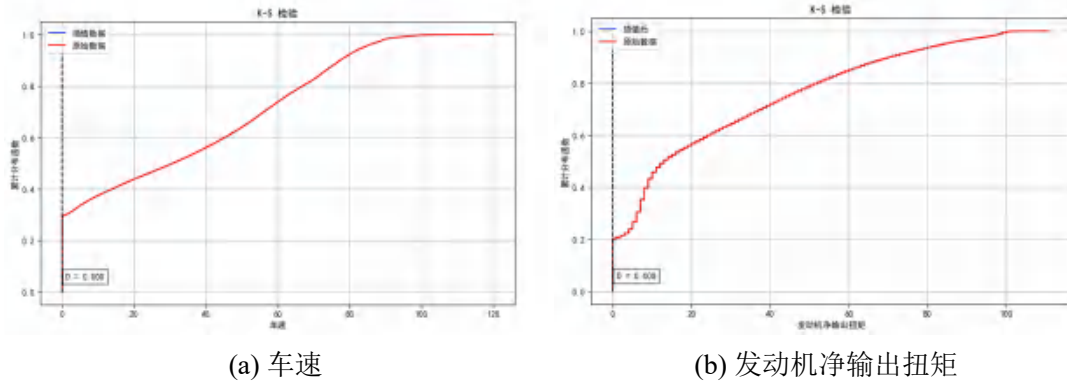


图 3.6 K-S 检验填补结果举例

从数据集中选取 13 个主要参数，包括车速、大气压力、发动机净输出扭矩、发动机摩擦扭矩等。使用以岭回归模型为估计器的递归特征消除方法（RFE），按照对 NO_x 传感器输出值的重要性，对参数进行排名，获取前五个最重要的特征，如下方 F_{RFE} 所示。

$$F_{RFE} = \{\text{speed, neot, eft, efflow, DPF}\}. \quad (3.2)$$

同时，计算 13 个主要参数与 NO_x 传感器输出值的皮尔逊（Pearson）和斯皮尔曼（Spearman）相关系数，并以热力图的形式对分析结果进行直观展示，如图 3.7、3.8 所示。在热力图中，颜色的深浅程度与特征值之间的相关程度呈正相关，即颜色越深表明相关程度越高，相关系数的绝对值越大；同时，当相关系数小于 0 时，表明两个特征值之间呈现负相关关系。根据相关系数综合分析可知，NO_x 传感器输出值与 neot、efflow、airin 以及 scrTin 具有较大的相关系数，说明了在一定程度上它们之间存在线性相关性单调相关关系。

由于 RFE 的估计器在一定程度上会影响参数的重要性排名，因此，将以上多维度的相关性分析以及 SCR 系统工作原理综合分析，选取如下 F_{corr} 所示的七个参数作为关联特征，用于建立多变量插值方法，填充 NO_x 传感器的缺失数据。

$$F_{corr} = \{\text{speed, neot, eft, DPF, efflow, airin, scrTin}\}. \quad (3.3)$$

(2) MLP-RF 模型架构设计

由于 SCR 系统 NO_x 传感器输出值受多方面因素的影响，这些影响因素之间也有较为复杂的耦合关系。因此，这里通过融合随机森林（RF）的特征交互优势与多层感知机（MLP）的非线性表达能力，建立用于实现 NO_x 传感器输出值缺失填补的模型。图 3.9 展示了 MLP-RF 模型的简要构建过程。具体架构设计流程如下：

首先，将基于多维相关性分析获取的 7 个参数作为输入特征（输入特征维度为 $R^{n \times 7}$ ），通过随机森林进行特征重要性排序，选择前 6 个关键特征（ X_{selected} ）



图 3.7 Pearson 相关系数热力图



图 3.8 Spearman 相关系数热力图

作为最终的输入特征，以进一步降低冗余信息对模型性能的干扰。其次，以优化后的随机森林模型作为基预测器，生成初步预测结果 $\hat{y}_{RF}$ 。最后，针对随机森林在复杂非线性关系建模上的局限性，构建 MLP 网络对预测残差进行学习，通过全连接层与 ReLU 激活函数的堆叠，自适应捕捉随机森林未建模的误差模式。最终的混合预测表达式如公式 (3.4) 所示。

$$\hat{y}_{\text{hybrid}} = \hat{y}_{RF} + f_{\text{MLP}}(X_{\text{selected}}). \quad (3.4)$$

该设计形成了优势互补的混合预测框架，既发挥随机森林的特征筛选能力，又借助 MLP 强化非线性关系建模，提升缺失值填补的精度。

(3) 特征工程

在 MLP-RF 模型建立过程中，对于输入数据的处理包括以下三个关键环节：其一，采用 Yeo-Johnson 变换处理输入数据。作为 Box-Cox 变换的扩展，Yeo-

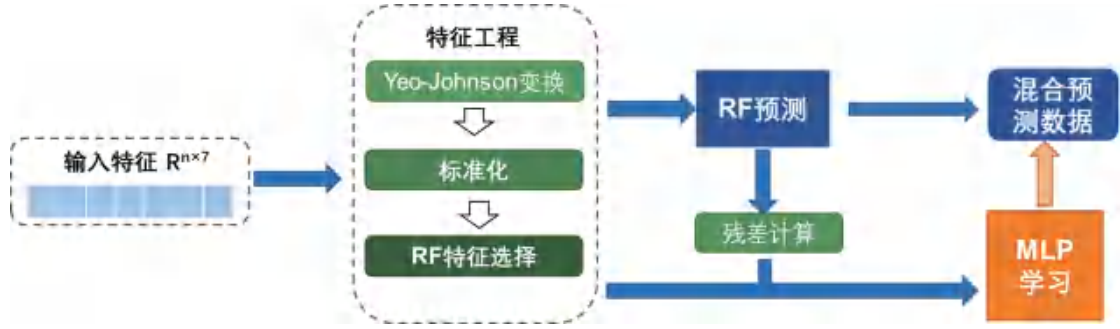


图 3.9 MLP-RF 模型简要构建过程

Johnson 变换通过优化参数 λ ，将非正态分布数据转换为近似正态分布形态，有效解决了 Box-Cox 变换仅适用于正数数据的局限性，防止特征偏态。其数学原理如公式 (3.5) 所示，式子表示当 $x \geq 0$ 时，通过幂函数或者对数函数压缩数据长尾，而在 $x < 0$ 时，通过特殊幂函数调整负数值的分布，最终让整体数据更接近正态分布，增强 MLP 对非线性关系的捕捉能力，提升模型的学习效率。

$$\begin{cases} \frac{(x+1)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, x \geq 0, \\ \ln(x+1) & \text{if } \lambda = 0, x \geq 0, \\ \frac{(1-x)^{2-\lambda} - 1}{\lambda - 2} & \text{if } x < 0. \end{cases} \quad (3.5)$$

其二，运用 Z-score 标准化方法对数据进行规范化处理，具体计算公式如 (3.6) 所示。式子中 μ 为特征均值， σ 为特征标准差。通过标准化操作，使特征均值归为 0、标准差归为 1，消除不同特征量纲差异对模型训练的干扰，确保算法对各特征的学习权重更合理。

$$z = \frac{x - \mu}{\sigma}. \quad (3.6)$$

其三，基于随机森林 (RF) 的特征重要性分析能力，筛选前六个重要特征作为模型输入参数。如图 3.10 所示，通过平均绝对 SHAP 值可视化呈现输入特征对上下游 NO_x 传感器输出值预测的影响程度。尽管上下游预测采用相同输入特征集，但各特征对上下游 NO_x 传感器输出值的影响权重呈现显著差异，该筛选过程进一步降低了冗余特征对模型性能的干扰。

(4) 参数优化

针对 MLP-RF 模型，构建分阶段参数优化框架，采用网格搜索，结合交叉验证方法，系统性探寻最优参数组合。其中，RF 与 MLP 的参数取值范围以及最佳取值结果分别如表 3.3、3.4 所示。

(5) MLP-RF 模型结果与对比

如图 3.11 所示，MLP-RF 模型对 SCR 系统上下游 NO_x 传感器输出值的预测表现优异，上游 R^2 达 0.915，下游 R^2 达 0.965，散点分布紧密围绕理想预测线，

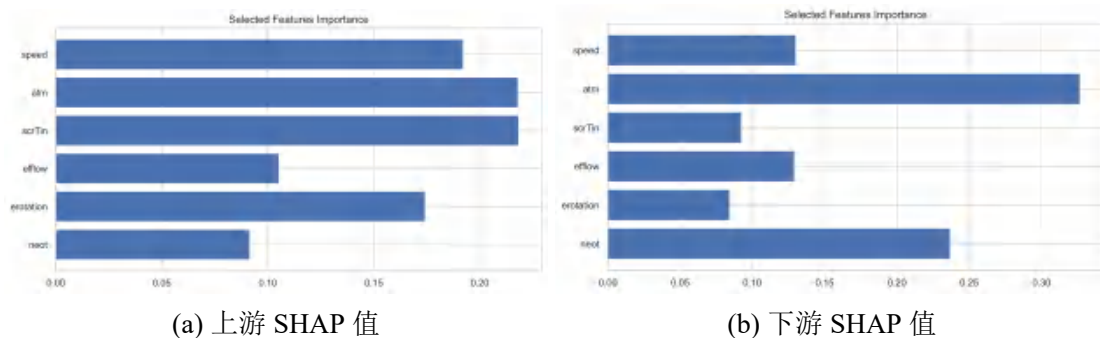


图 3.10 MLP-RF 模型特征重要性分析

表 3.3 RF 参数取值

参数	取值范围	最佳取值
n_estimators	[100,200,300]	200
max_depth	[10,15,20]	15
min_samples_split	[2,4]	4
max_features	['sqrt', 0.7]	'sqrt'

表 3.4 MLP 参数取值

参数	取值范围	最佳取值
hidden_layer_sizes	[(64,34),(64,32,16)]	(64,32,16)
α	[0.001,0.0001]	0.001
batch_size	[128,256]	256

表明模型对上下游数据均具有强拟合能力。结合表3.5的模型对比与消融实验结果：相较于 SVR、KNN、RF、MLP 等基准模型，MLP-RF 模型在上游预测中 R^2 提升最高达 0.227（对比 SVR），下游预测 R^2 提升最高达 0.313（对比 SVR），尤其对下游 NO_x 传感器输出值的预测精度提升显著，这对精准获取 NO_x 排放数据具有关键意义。因此，MLP-RF 模型可有效应用于 NO_x 传感器输出值的插值任务。

通过上述多阶段数据预处理（含特征工程、参数优化等环节），最终获取国六重型柴油货车真实运行环境下的高质量 OBD 数据集，为后续排放因子计算与分析奠定坚实数据基础。

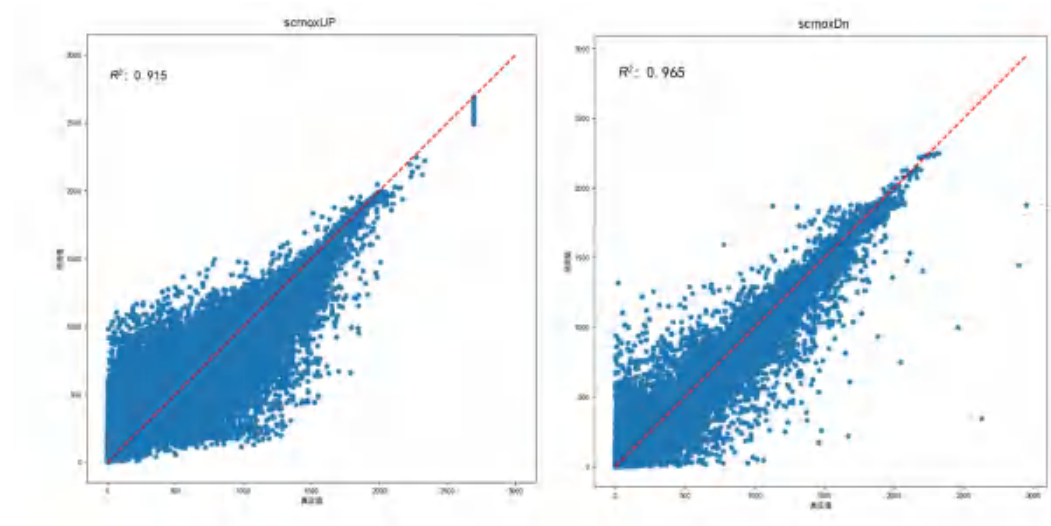


图 3.11 MLP-RF 模型真实-预测值

表 3.5 模型结果对比

模型	R^2 (上游)	R^2 (下游)
SVR	0.688	0.652
KNN	0.860	0.849
RF	0.867	0.858
MLP	0.824	0.815
MLP-RF	0.915	0.965

3.2 工况数据

3.2.1 车辆比功率 VSP

车辆比功率 (Vehicle Specific Power, VSP) 是由 Jimenez 等^[37]于 1999 年首次提出, 定义为车辆瞬时功率和机动车质量的比值 (单位: kW/t)。该指标主要目的是量化发动机在克服行驶阻力 (包括滚动摩擦阻力、空气阻力等), 提升车辆动能与势能时需要的输出功率。作为车辆排放评估的一个综合性能指标, VSP 能够有效关联车辆运行状态与尾气排放特征。大量应用研究表明, 车辆的尾气排放与 VSP 之间存在显著相关性, 这种关联特性使得 VSP 在车辆排放研究中得到了广泛应用, 如车辆尾气污染物排放因子的计算等。基于此, MOVES、IVE 等多种模型将 VSP 作为关键参数, 借助其与排放之间的强相关性, 分析 VSP 与尾气中各种污染物排放之间的关系, 实现对不同工况下车辆排放的计算与模拟, 来更深入地理解车辆排放的内在机制, 为建立准确的排放因子模型提供依据。

在不同模型中, VSP 的计算方式存在一定的差异。以 MOVES 模型为例, 其 VSP 计算方法基于丰富可靠的基础数据集获取, 具有较强的灵活性和可扩展性,

计算过程综合考虑了车速、加速度、质量、车辆滚动阻力系数等参数，且覆盖重型货车各类运行场景；而 IVE 模型则是根据发动机功率、车速、质量来确定 VSP。相对而言，IVE 在基础数据集的完备性、参数考量的综合性，以及对重型货车多样化运行场景的适配能力上，均不及 MOVES 模型。

在行驶过程中，重型柴油货车的运行状态与排放不仅仅受路况、速度、加速度等运行参数的影响，其自身的质量也会改变影响发动机的负荷。因此，为了获取更加准确的工况数据，本研究综合考虑各种影响因素，选用 MOVES 模型中的 VSP 计算方法，将车重也考虑其中，并将选用的重型柴油货车按照其总质量划分为五种类型。根据 MOVES 模型中 VSP 的计算方法，依据不同的参数，获得车辆的 VSP 数据。具体的 VSP 计算公式如式 (3.7) 所示。

$$VSP = \frac{Av + Bu^2 + Cv^3 + mva}{f_{scale}}, \quad (3.7)$$

其中， v 为瞬时速度，单位 m/s； a 为加速度，单位 m/s^2 ； m 为车辆及负载总质量，单位 t（吨）； A 为车辆滚动阻力系数； B 为旋转滚动阻力系数； C 为空气阻力系数； f_{scale} 为换算系数。对于 A 、 B 、 C 、 m 以及 f_{scale} 五个系数，按照车辆吨位进行取值，具体取值如表 3.6 所示。

表 3.6 VSP 计算公式系数取值

车辆类型	A	B	C	m	f_{scale}
重型货车 (12-16t)	0.9872	0	0.003616	14	17.1
重型货车 (16-22t)	1.3002	0	0.004456	19	17.1
重型货车 (22-28t)	1.6758	0	0.005200	25	17.1
重型货车 (28-40t)	2.2392	0	0.005776	34	17.1
重型货车 (40t 以上)	2.6148	0	0.005800	40	17.1

3.2.2 VSP 区间划分

通过 VSP 公式计算获取车辆的 VSP 工况值，为了更加详细的分析 VSP 数据，选取合适的 VSP 区间（即 VSP Bin）划分方案，对车辆工况数据进行聚类分析。本次研究中 VSP 区间的划分方法依旧结合了 MOVES 模型中提供的聚类方式，并依据重型柴油货车的运行数据特征进行了部分改进。首先根据加速度与速度数据，将数据划分。其中包括两个加速度区间，即 $a_t \leq -0.9m/s^2$ 与 $a_t > -0.9m/s^2$ ，以及三个速度区间，即 $[-1.6, 1.6)km/h$ 、 $[1.6, 50)km/h$ 、 $[50, 80)km/h$ 和 $[80, +\infty)km/h$ ，然后进一步结合 VSP 值所处的 VSP 区间，对车辆运行状态加以定义，包括减速、怠速、滑行、加速四种运行状态。最终共划分了 29 个 VSP 区间，具体的划分方法如表 3.7 所示。

通过 VSP 计算获取车辆运行工况，利用 VSP 区间划分对其运行状态添加标

签 VSP Bin，并将这些数据汇总到 OBD 数据集中，用于车辆的运行工况分析。

表 3.7 重型货车 VSP 区间划分

VSP Bin ID	VSP 区间	运行状态描述	机动车速度 $v(\text{km/h})$	机动车加速度 $a_t(\text{m/s}^2)$	
0	-	减速/刹车		$a_t \leq -0.9$	
1	-	怠速	$[-1.6, 1.6)$		
11	$(-\infty, 0)$	滑行			
12	$[0, 1)$				
13	$[1, 2)$				
14	$[2, 3)$				
15	$[3, 4)$	巡航/加速	$[1.6, 50)$		
16	$[4, 5)$				
17	$[5, 6)$				
18	$[6, 9)$				
19	$[9, +\infty)$				
21	$(-\infty, 0)$	滑行			
22	$[0, 1)$				
23	$[1, 2)$				
24	$[2, 3)$			$a_t > -0.9$	
25	$[3, 4)$	巡航/加速	$[50, 80)$		
26	$[4, 5)$				
27	$[5, 6)$				
28	$[6, 9)$				
29	$[9, +\infty)$				
31	$(-\infty, 0)$				
32	$[0, 1)$				
33	$[1, 2)$				
34	$[2, 3)$				
35	$[3, 4)$	巡航/加速	$[80, +\infty)$		
36	$[4, 5)$				
37	$[5, 6)$				
38	$[6, 9)$				
39	$[9, +\infty)$				

3.3 道路信息数据

3.3.1 车辆道路定位

在对车辆运行状态分析的过程中，获取车辆运行所在道路的信息也是重要的一步。考虑到需求和硬件资源，本次研究基于数据集中的经纬度数据，使用全球行政边界数据集（Global Administrative Areas, GADM）与开放街道地图（Open Street Map, OSM）资源，实现从省级地图到区域路网的逐级定位，获取车辆运行所在的道路，并成功提取道路编号以及道路类型等关键信息。该过程的主要流程

如图3.12所示，具体步骤如下：

首先，获取准确的省级行政区范围数据。GADM 是一个广泛应用于地理信息研究的数据库，提供了详细的全球行政区划信息，能够为车辆省级定位提供基础数据。因此，从 GADM 获取中国地图省级行政区划的离线数据，并使用 QGIS 软件对这些数据集进行可视化，以核对省级行政区划图。通过检查属性表中的省份代码、省份名称等客观信息，修正省份名称与地图板块对应问题，确保了地图中的各省份板块的经纬度信息与实际行政区的经纬度范围一致性，并据此成功获取了 34 个省级行政区的多边形经纬度边界，为车辆的省级定位提供了基础。

其次，获取每辆车的运行范围，确定车辆运行路径所属的省份。根据车辆的 VIN 码将 OBD 数据划分为多辆车的子数据集。利用子数据集，获取每辆车运行轨迹的经纬度范围，建立有效的经纬度边界框，利用 Python 的 geopandas 数据处理，与省级行政区的多边形经纬度边界匹配，获取车辆路径所在的省份，并使用 geojson 进行数据存储。

然后，获取各省份的路网数据。OSM 数据集是一个开源可编辑的地理数据集，提供不同划分等级的路网数据，其中的道路数据包含了高速公路、国道、省道、县道等多类道路信息，并且对不同道路类型都有道路 ID 和道路类型（highway）等多种属性，如高速公路是 highway= ‘motorway’，国道是 highway= ‘trunk’ 等。从 OSM 获取我国 34 个省份的路网数据，并结合车辆的经纬度边界框获取车辆运行区域的路网图，以确保覆盖车辆的整个运行区域。

最终，利用经纬度信息，结合已获得的路网地图，精确定位车辆所在的具体道路，并提取道路信息中的 highway 和道路 ID 属性，将这些关键信息加入到当前数据集中。

在获取道路数据的过程中，因为考虑到硬件资源有限以及计算效率，所以采取从省份到区域的逐级道路定位。通过以上操作，可以成功提取车辆运行所在的道路信息，获取道路类型，为后续的数据的分类分析和应用提供了坚实的基础。

3.3.2 道路分类

在 OSM 数据集的道路分类体系中，对道路进行了多种分类，包括高速公路（motorway）、国道（trunk）、主要道路（primary）、次要道路（secondary）、未分类道路（unclassified）等 11 个层级。根据我国道路等级划分规范^[38]，我国的道路等级分为快速路、主干路、次支路和支路四类。以 OSM 提供的道路类型作为子类，将其道路类型与我国的道路等级分类结合，以车辆运行速度限制为参考，共归并为三个大类，即快速路（Motorway）、主干路（Primary）以及次支路（Secondary），每种类型包含的具体子道路类型如表3.8所示。分类后，既可以保

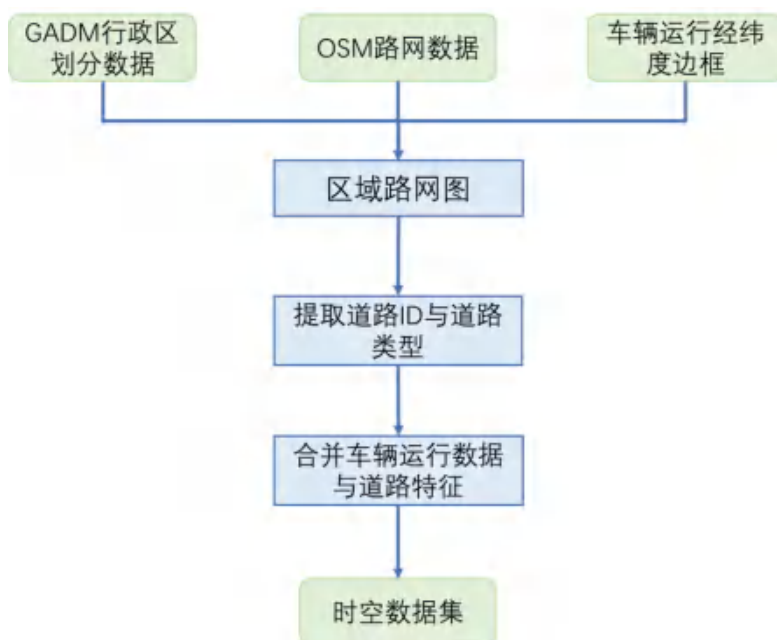


图 3.12 道路信息定位流程图

留 OSM 数据的多元特征，有确保了与我国道路特征的兼容，为后续车辆道路运行工况分析提供统一规范的分类基础。

表 3.8 道路分类

道路类型	子道路类型
Motorway	motorway、motorway_link、trunk、trunk_link
Primary	primary、primary_link
Secondary	secondary、unclassified 等其余道路类型

3.3.3 车辆行程划分

在长时间的连续运行过程中，车辆的工况组成较为复杂。若把一个长时间运行数据作为一个整体进行分析，无法很好的获取其中的细节数据，难以观察工况分布。因此，基于 OBD 数据集以及道路信息数据，获取了车辆运行所在的道路信息与时间数据，对车辆的行驶轨迹进行分割。

已有研究表明^[39]，当车辆的轨迹分割时间长度介于 60s 至 300s 之间时，运行信息的缺失率以及相对误差均可以维持在 2% 以下，是较为良好的时间窗口长度选取范围。依据重型柴油货车的行驶特点，本次实验选取 180s 为基准时间窗口，既可以较好地反应车辆在不同行驶工况下的阶段性特征，又不会让划分后的数据过度碎片化。另外，考虑到快速路上车辆行驶状态相对稳定且车速一般较高，为了可以更细致地捕捉变化，选取 60s 的间隔对快速路行程进一步划分。对车辆运行数据进行细化处理，获取更便于分析的数据集单元。这不仅能够有效降

低数据处理的复杂性，还能有效保留行程中的关键道路信息，为后续车辆运行特征分析、尾气排放研究等工作奠定基础。

行程划分流程如图3.13所示，具体操作主要包括以下步骤：

(1) 基准时间窗口划分行程：依据车辆 VIN 码和数据采集时间，将数据集进行升序排列，然后采用 180s 的基准时间窗口对单辆车的连续行程轨迹进行初步划分。

(2) 行程道路属性确定：对于每个划分出的行程段，确定其行驶的道路类型和对应的道路 ID。对于每个行程段，判断其首、中、尾三部分中是否有道路的占比达到或超过 2/3。如果是，记录该道路的信息以及对应的行程编号，判断其所属的道路类型，并将这些信息作为行程的道路属性加入数据，为了解行程段所在的道路提供数据，为后续数据集的进一步划分提供依据。

(3) 快速路行程轨迹划分：判断行程的道路属性类型，若属于快速路，则使用 60s 的时间窗口进行进一步细分，并更新行程编号。

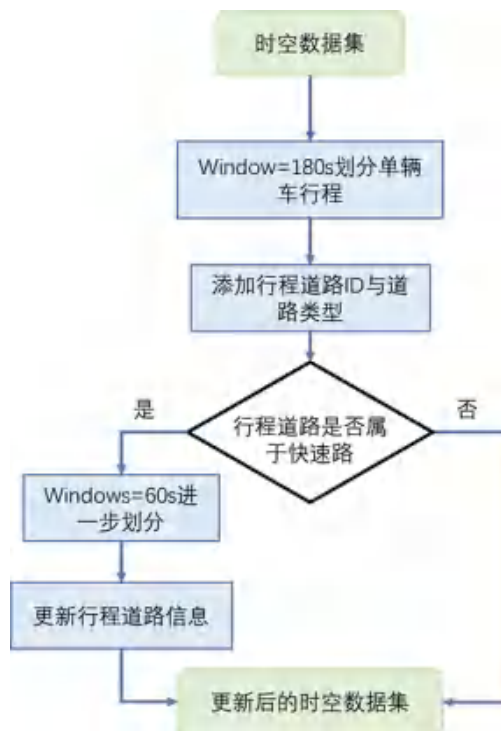


图 3.13 车辆行程划分流程图

3.3.4 车辆运行工况分析

基于 3.1 节至 3.3 节的数据处理与融合工作，构建了包含车辆运行数据、工况数据以及道路信息的重型柴油货车时空数据集。为了分析车辆运行工况分布的合理性，结合车辆类型编号与行程道路类型作为标签进行数据集划分，获取在三种道路类型上，五种车辆类型各自的数据集，并以 VSP 区间（VSP Bin）为工

况参考，综合车辆速度、道路、质量三个特征，查看车辆运行工况数据的分布。

首先，求取一辆车每个行程的平均速度，按照行程平均速度所属的速度区间，添加每条运行数据的速度标签（speed_label）。然后，先以道路类型和速度标签作为分类依据对数据进行分组，接着按照车辆类型编号对数据进一步划分，统计在每个速度区间内，不同类型的车辆在不同道路类型上的 VSP 工况分布特征。这里选取了其中四个具有代表性的速度区间（24-26km/h、34-36km/h、44-46km/h、54-56km/h）的 VSP Bin 占比柱状图进行展示，如图3.14至3.16所示。

结合 VSP 计算方法与 VSP Bin 获取方法，分析多个平均速度区间内各种 VSP 工况数据分布可知，每种类型的车辆在不同工况下分布合理，符合重型柴油货车速度-VSP 工况的变化规律^[40]，且数据分布均匀，用于 NO_x 排放因子计算和 NO_x 排放建模时不会因为某一情况聚集而导致结果偏差。

以列举图像为例，在低速区间 24-26km/h 中，五种车辆类型的 VSP Bin 都集中分布在低速对应的 VSP 区间，这是由于在低速行驶时，车辆的动力需求较低，发动机的工作负荷相对较小，VSP 值对应也处于较低水平。随着行程平均速度的提高，速度所属区间向高速区间变化的同时，代表 VSP Bin 占比的柱子集中向右移动，VSP 也在升高，这一变化符合 VSP 与车速变化的规律。当车辆速度提高时，发动机需要输出更多的功率来克服空气阻力和行驶阻力，从而使得 VSP 值相应增大，VSP Bin 分布也随之向高速区间偏移。这不仅验证了 VSP 作为衡量车辆运行工况的有效性，也可以为进一步研究车辆的排放特性提供了重要的数据依据。

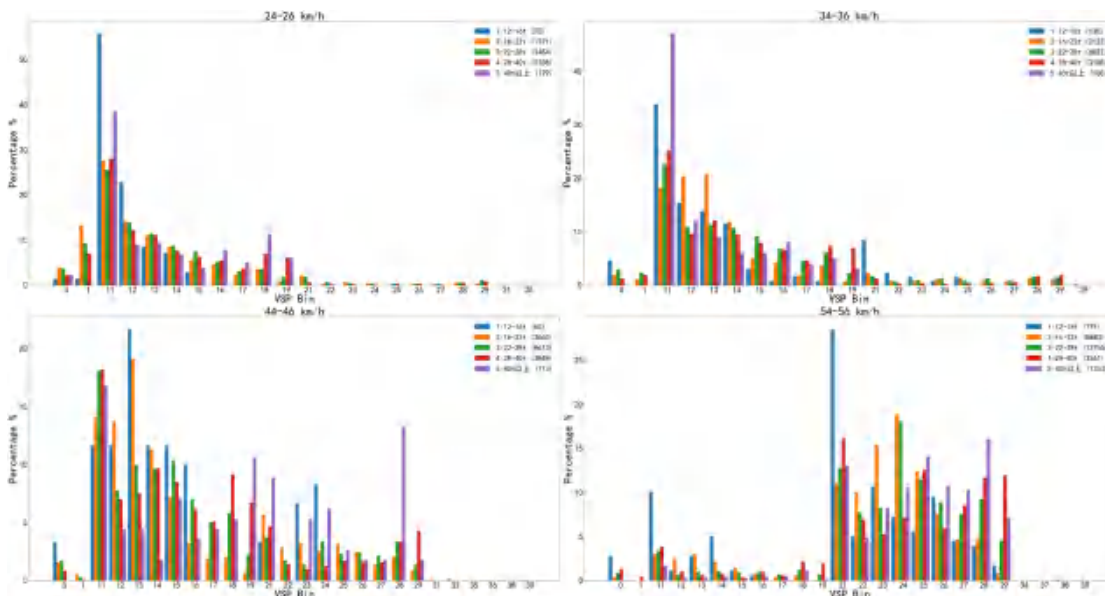


图 3.14 Motorway 上特定速度区间下 VSP Bin 分布

综上所述，通过数据处理与融合，耦合驾驶工况与道路特征，建立了重型柴油货车在真实道路上的时空数据集，且分析了该数据集涵盖车辆工况的全面性

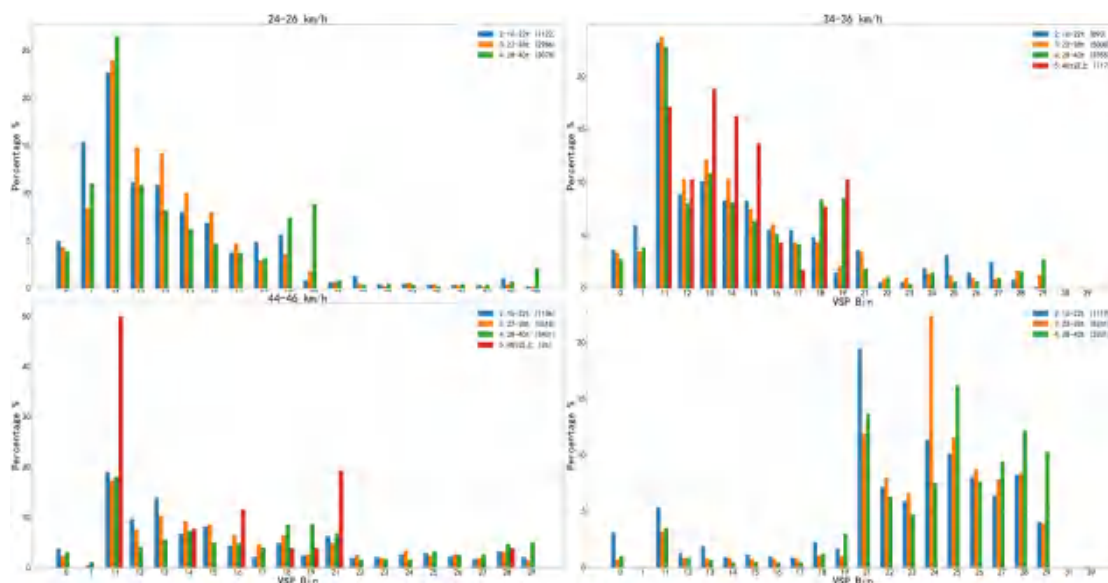


图 3.15 Primary 上特定速度区间下 VSP Bin 分布

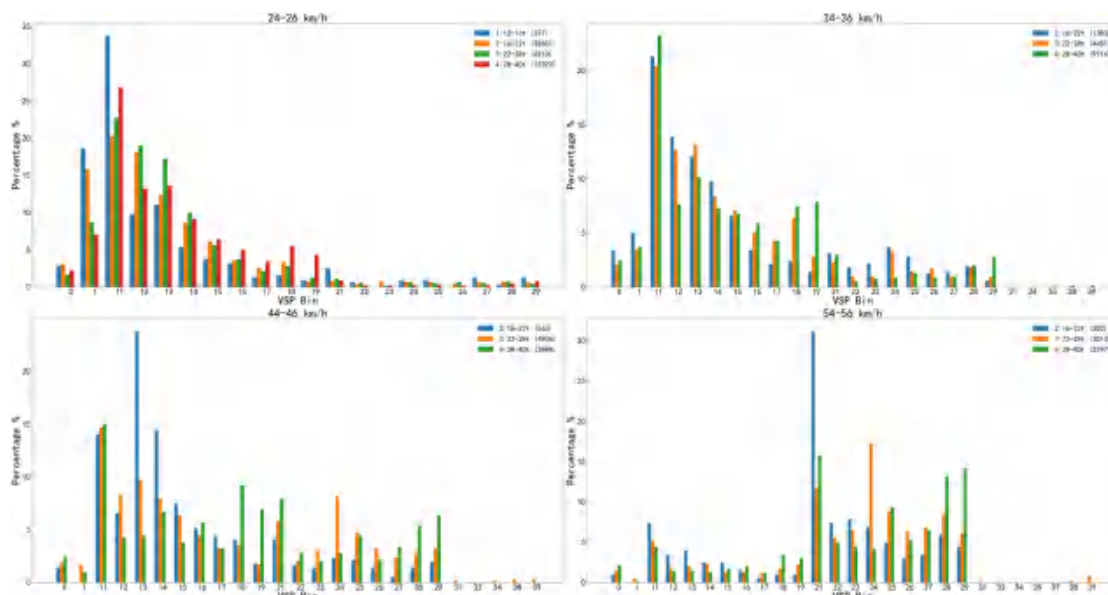
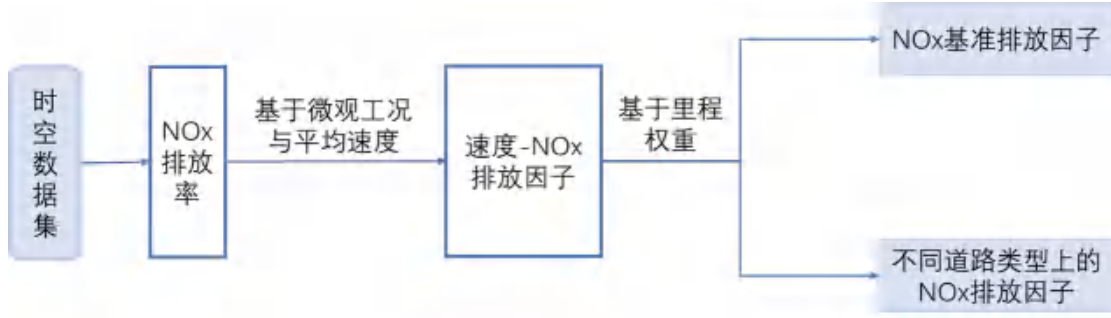


图 3.16 Secondary 上特定速度区间下 VSP Bin 分布

与合理性。该数据集为获取国六重型柴油货车的 NO_x 排放因子以及 NO_x 近实时瞬态排放模型的构建奠定了良好的数据基础。

3.4 NO_x 排放因子

基于时空数据获取 NO_x 的排放因子，需要先获取 NO_x 排放率以及不同速度下的 NO_x 排放因子数据，结合一定的权重加和，获取最终的 NO_x 排放因子，简要的过程如图3.17所示，具体步骤则见下方小节内容。

图 3.17 NO_x 排放因子获取过程

3.4.1 NO_x 排放率

1. NO_x 排放率计算

重型柴油货车尾气污染物中的 NO_x 主要源于发动机燃烧过程。从 NO_x 的产生原理来看^[41]，进气量和喷油量对 NO_x 的生成有着至关重要的影响。进气量决定了燃烧过程中氧气的供应量，在高温环境下，充足的氧气会促使空气中的氮气与氧气发生反应，生成热力 NO_x。喷油量则直接关系到参与燃烧的燃料量，并且车辆加速、高负荷等状态下，会存在局部燃料过浓的区域，此时碳氢自由基与氮气反应会导致快速 NO_x 的产生，同时燃料中含有的氮元素也会在燃烧时转化为燃料 NO_x。

基于上述 NO_x 产生的原理，结合所使用的车辆运行数据集，本次实验采用基于远程监控数据构建的 NO_x 排放率计算公式，综合考虑进气流量、喷油量以及 SCR 下游 NO_x 传感器输出值对 NO_x 排放结果的影响，计算公式如 (3.8) 所示。

$$E_{NO_x} = \left(\frac{\text{进气流量}}{3600} + \frac{\rho \times \text{喷油量}}{3600} \right) \times D_{NO_x} \times 0.001587, \quad (3.8)$$

其中， E_{NO_x} 为 NO_x 排放率，单位是 g/s。进气流量为传感器进气量输出值，单位是 kg/h， ρ 为柴油密度，取值为 0.84kg/L，喷油量为车辆的燃料流量，单位是 L/h， D_{NO_x} 浓度为车辆的 SCR 下游 NO_x 传感器记录的 NO_x 气体浓度，单位是百万分之一 (ppm)。

2. NO_x 排放率变化特征

按照车辆吨位这个特征值，将重型柴油货车细分为 5 个吨位范围，绘制每个吨位范围的 VSP Bin-NO_x 排放率箱型图，分析 NO_x 排放与车辆工况之间的变化规律。

结合图3.18可知，不同吨位类型的重型柴油货车在运行过程中，整体上都体现出的规律是：将一个速度范围作为一个整体，则在不同的速度范围之间，随着 VSP Bin 增大，速度区间内车辆的 NO_x 排放率整体上比上一个速度区间降低；在同一个速度区间中，随着 VSP Bin 增大，VSP 升高，车辆的 NO_x 排放率也升高。结合以往的研究结论可知^[42-43]，这里的重型柴油货车在微观工况下的 NO_x 排放

率分布符合已有研究结论与实际情况，这也从侧面说明了本次所构建的数据集的真实性与合理性。

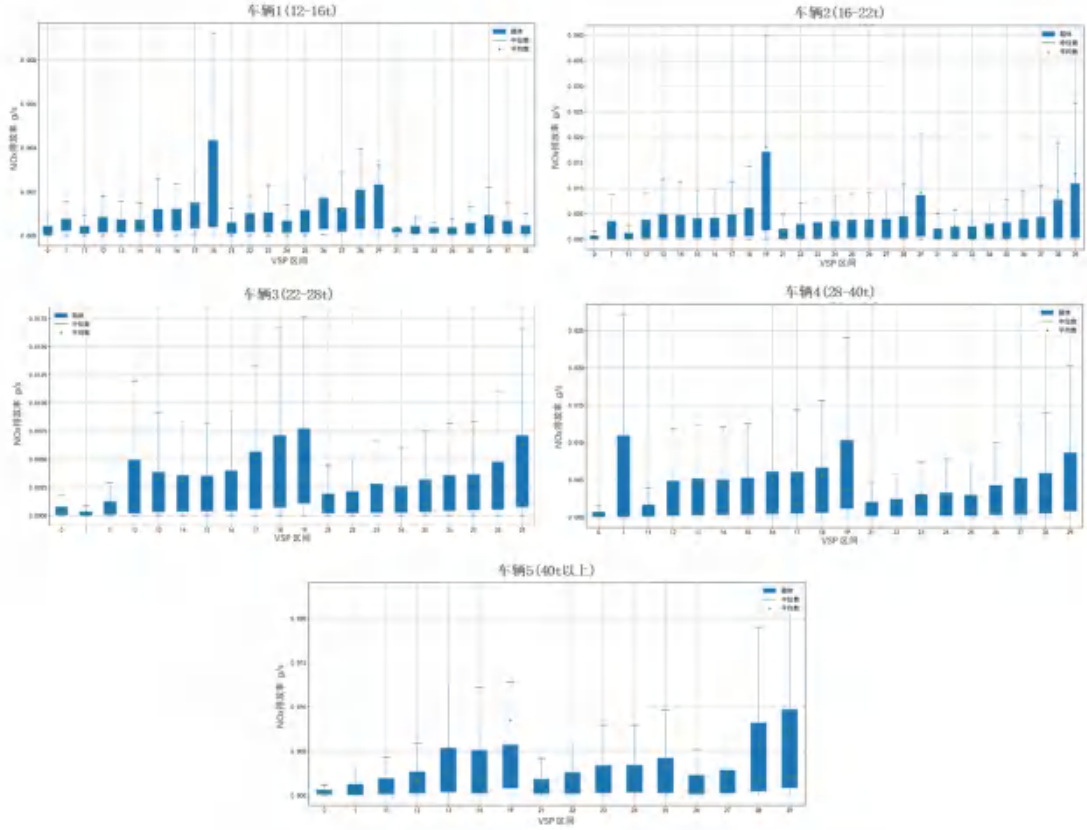


图 3.18 VSP 区间下的 NO_x 排放数据分布

3.4.2 速度-NO_x 排放因子

利用 NO_x 排放率，结合行程数据与道路信息，求取每种道路类型下，NO_x 排放因子随速度变化的趋势。

1. 划分数据集

首先，将五种吨位类型的车辆数据划分成用于计算 NO_x 排放因子的数据集，包括两个划分标签。一是结合道路类型将每种车辆类型进一步分组，形成快速路（Motorway）、主干路（Primary）、次支路（Secondary）三组数据。二是根据之前行程划分后获取的行程平均速度标签，将每种车辆类型的三个道路数据集中的行程数据按照平均速度标签汇总，每个速度标签对应一个新的道路-平均速度-车辆类型子集（以下简称速度子集）。比如车辆类型 1 的某个快速路行程的平均速度位于 [0,2)km/h 的区间，该区间平均速度标签为 1km/h，那么该行程数据就归入（快速路-1km/h-车辆类型 1）子集中。

2. 计算平均速度对应的 NO_x 排放因子

在速度子集中，首先，利用计算公式 (3.9) 计算子集中，每个微观工况区间的平均 NO_x 排放率，然后以子集中 VSP Bin 的占比为权重，子集对应的平均速

度为车速，利用公式 (3.10)，获得当前子集的 NO_x 排放因子，即平均速度对应的 NO_x 排放因子。

$$\overline{ER}_{i,j} = \frac{1}{T_{i,j}} \sum_{t=1}^{T_{i,j}} ER_{i,j,t}, \quad (3.9)$$

$$EF_{i,s} = \frac{3600 \times \sum_{j=Bin0}^{Bin39} (\overline{ER}_{i,j} \cdot P_j)}{\bar{v}_s}, \quad (3.10)$$

其中， i 代表车辆类型， j 代表工况区间 VSP Bin，用来表示微观工况 VSP， $T_{i,j}$ 为车辆类型 i 在工况区间 j 的总时间，单位为 s。ER_{*i,j,t*} 为车辆 i 在工况区间 j 下第 t 秒的排放速率，单位为 g/s。ER_{*i,j*} 是车辆 i 在 Bin j 下的平均 NO_x 排放率，单位为 g/s， P_j 为 Bin j 在当前子集中的占比， $\bar{v}_s$ 是当前速度子集 s 对应的平均速度，单位为 km/h，EF_{*i,s*} 即为车辆 i 在当前速度 $\bar{v}_s$ 下的 NO_x 排放因子，单位为 g/km。

3. 平均速度-NO_x 排放因子曲线图

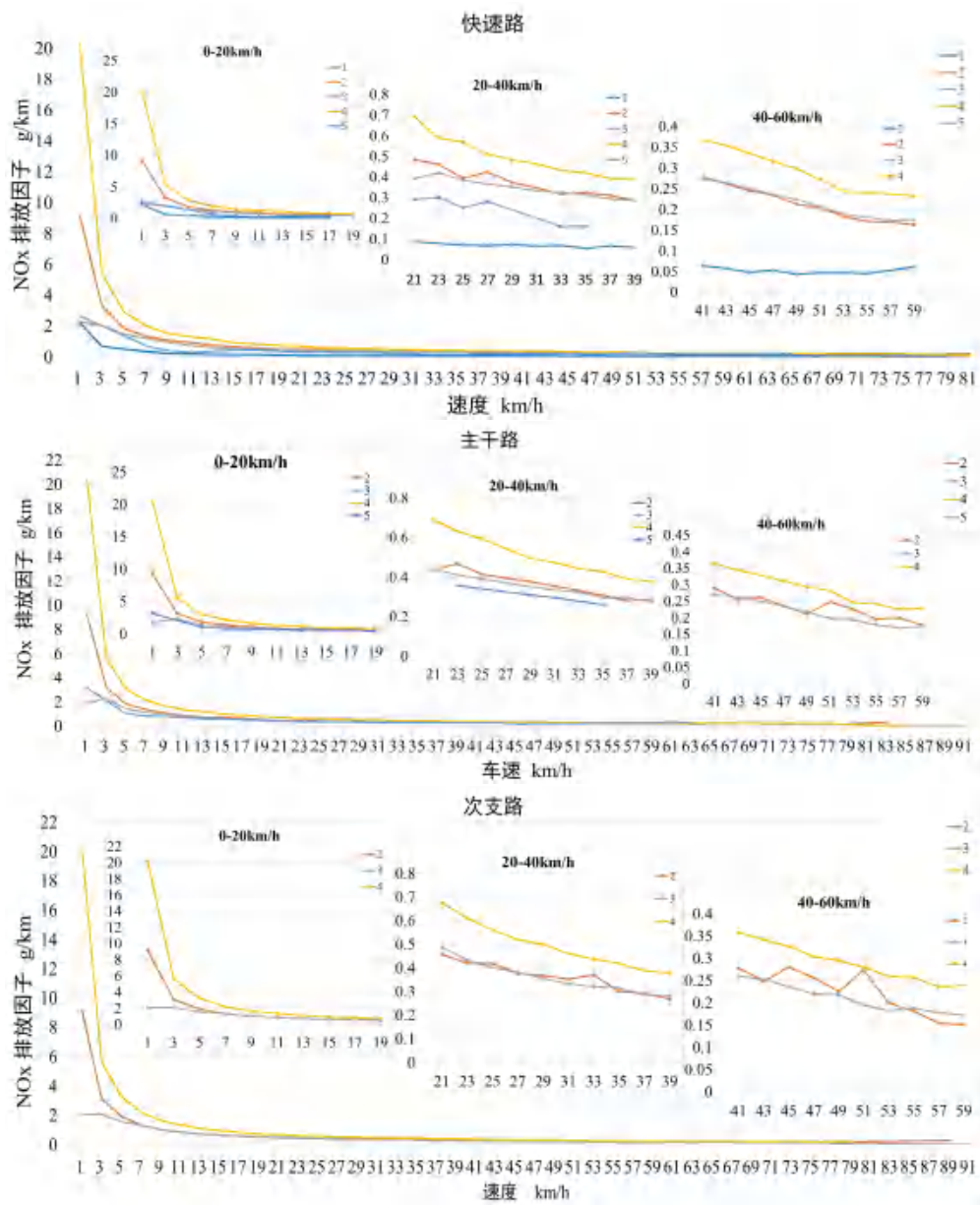
对全部速度子集计算后，便获取每种车辆类型在不同平均速度下的 NO_x 排放因子，单位为 g/km。不同车辆类型在不同道路上的速度-NO_x 排放因子的结果如图3.19所示。整体上，不同车辆类型的 NO_x 排放因子都是随着车辆运行速度的提高而随之降低，并逐渐趋于平缓。

结合中国机动车排放模型以及其他相关研究中对 NO_x 排放因子变化趋势的研究结论可知^[44-45]，图3.19的 NO_x 排放因子曲线变化趋势符合真实情况。具体变化原因为：在 0-10km/h 的低速工况下，由于重型柴油货车在低速运行时，发动机转速偏低，燃油喷射系统无法将燃油与空气进行充分的混合，燃油无法完全燃烧，导致车辆 NO_x 排放量较高。另外，低速工况下，重型柴油货车的 SCR 尾气后处理系统中催化转化器的工作温度一般较低，无法达到可以使催化剂实现有效转化的温度，这同样会导致 NO_x 排放量高，因而导致该速度工况下的 NO_x 排放因子较高。

在速度达到 20km/h 左右时，车辆的 NO_x 排放因子明显降低。这是由于车速增加的同时，发动机转速也随之提升，这使得燃油与空气能够更好地混合，燃烧更加充分，同时，尾气温度的升高也会提高 SCR 尾气处理装置的入口温度，从而提升催化剂的转化效率，减少 NO_x 的排放，降低 NO_x 排放因子。

而当速度大于 20km/h 后，车辆的发动机进入相对稳定的工作状态，有较高的燃烧效率，且尾气后处理系统中催化器的工作温度合适，可以实现对 NO_x 的高效率转化，从而使 NO_x 排放维持在较低的水平，NO_x 排放因子随速度的变化也趋于平缓。

结合相关研究与以上分析可得，本实验获取的 NO_x 排放因子的分布以及变化规律符合重型柴油货车在实际运行中应有的变化规律，是合理的分布结果，说明了本次数据的合理性与结果的真实性。

图 3.19 NO_x 排放因子曲线图

3.4.3 重型柴油货车 NO_x 排放因子

1. 不同道路类型上的综合 NO_x 排放因子

首先，基于上方计算出的平均速度-NO_x 排放因子，以每个速度子集的里程在一个车辆类型的总里程的占比为排放因子权重，加和求得每种车辆类型在不同道路上的 NO_x 排放因子。其中，在快速路上，车辆 1 到 5 类型的 NO_x 加权排放因子分别为：0.037g/km, 0.19g/km, 3.42g/km, 0.47g/km, 0.23g/km。主干路上，车辆类型 2 到 5 的 NO_x 加权排放因子分别为：0.34g/km, 1.76g/km, 0.54g/km, 0.48g/km，而车辆类型 1 由于原始数据集中不具备其在主干路上的数据，因此没有获取该类型的车辆在主干路上的综合排放因子。次支路上，车辆 1 到 5 类型的 NO_x 加权排放因子分别为：0.084g/km, 0.42g/km, 1.77g/km, 0.94g/km, 0.34g/km。

接着,利用每种车辆类型在一种道路类型上的里程占比为权重,计算三种道路上各自的综合排放因子,获取图3.20所示的快速路、主干路、次支路三种类型下的综合 NO_x 排放因子,其值分别为: 1.82g/km 、 1.28g/km 、 1.24g/km 。根据国六重型货车排放标准^[4]要求:国六标准下重型货车的 NO_x 排放比国五标准加严了 77%,结合《道路机动车大气污染物排放清单编制技术指南》^[46]中的重型柴油货车 NO_x 基准排放因子可知,国五标准下为 4.72g/km 。因此,可判断,本次实验获取的重型货车在三种道路类型上的综合 NO_x 排放因子均为正常值。

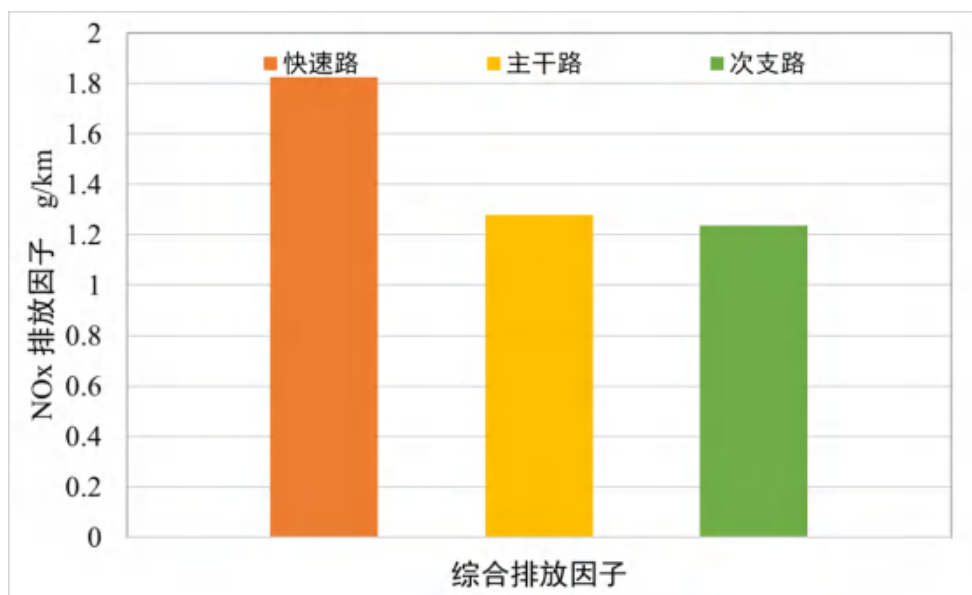


图 3.20 NO_x 综合排放因子

2. 基准 NO_x 排放因子

在建立宏观排放清单时,需要使用的时基准 NO_x 排放因子,无需详细的数据分类。因此,基于不同道路类型上的车辆里程在时空数据集总里程中的占比,对 NO_x 综合排放因子进行加权,获取基准 NO_x 排放因子为 1.64g/km ,相较于国五重型柴油货车的基准排放因子下降了约 65%,属于有效合理值。

基于上述过程,获取了国六重型柴油货车的排放因子,填补了真实环境下的 NO_x 国六排放因子的空白。进一步分析可知,国六标准的施行对重型柴油货车的 NO_x 减排效果显著,非常有利于 NO_x 污染管控工作的进行。另外,通过对比不同道路类型上的综合 NO_x 排放因子可知,快速路上的 NO_x 排放比主干路、次支路高出 29.7%-31.9%,由此可以看出道路特征对 NO_x 排放的影响,在之后的 NO_x 排放管控中应该相应的策略,进一步加强对快速路上重型柴油货车 NO_x 排放的精准管理工作。

3.5 本章小结

本章首先结合线性插值、多层感知机、随机森林算法，利用分阶段插值方法完善基于 OBD 获取车辆运行数据。其次利用 MOVES 模型中的 VSP 计算方法获取车辆的 VSP 工况数据以及 VSP Bin 标签。然后结合 GADM、OSM 以及经纬度定位获取车辆的道路信息，并与车辆运行工况数据融合，获取了耦合驾驶工况-道路特征的国六重型柴油货车时空数据集。基于时空数据集，通过对其运行工况分布以及工况- NO_x 排放变化特征进行可视化分析，获取重型柴油货车工况- NO_x 排放的变化特征，并结合以往研究判断分布的合理性。并以微观工况占比为权重，计算不同速度子集中的 NO_x 排放因子，分析重型柴油货车 NO_x 排放随速度的变化特征，其次是以里程占比为权重，获取了真实道路上重型柴油货车的 NO_x 基准排放因子为 1.64g/km，以及快速路、主干路、次支路三种类型下的综合 NO_x 排放因子分别为：1.82g/km、1.28g/km、1.24g/km，对比可知快速路较主干路和次支路的排放强度高 29.7%-31.9%，显示出道路特征结构对 NO_x 排放的关键影响。

第4章 基于堆叠泛化的 NO_x 瞬态排放预测方法

本章节针对传统排放模型没有考虑 NO_x 瞬态排放的问题，提出了一种基于堆叠泛化集成算法的重型柴油货车 NO_x 瞬态排放预测模型。本次研究基于第三章获取的重型柴油货车的时空数据集，首先借助相关系数分析 NO_x 瞬态排放与其他特征的相关性，筛选初始的输入特征；其次，利用 XGBoost 模型进行特征重要性评估，进一步结合特征消融，获取与 NO_x 瞬态排放最为相关的车辆运行特征，完成特征优化；接着，采用堆叠泛化集成方法，融合 LightGBM 和 XGBoost 两种基模型的优势，获取表现最佳的 NO_x 瞬态排放预测模型，即 SLXGB 模型。并与其他模型方法对比，验证本次建立的模型的有效性。最后，将 SLXGB- NO_x 瞬态排放预测模型用于建立成都市重型柴油货车高分辨率的 NO_x 排放清单，并分析 NO_x 排放的时空分布特征。

4.1 数据集与相关性分析

通过第三章的工作，本研究已经获取了耦合驾驶工况-道路特征的重型柴油货车逐秒级的时空数据集，其中包含符合国六标准的不同类型重型柴油货车在不同道路上的运行数据，地理范围上遍布全国多个省份。并利用公式 (3.8) 获取了 NO_x 的原始瞬时排放率。为了获取数据集中各特征与 NO_x 瞬时排放的相关性，这里依然选用 Pearson 和 Spearman 相关系数，分析 NO_x 瞬时排放与其他特征的相关性，如图4.1。但是，由于原始 NO_x 排放率数据的计算公式中使用了进气量、燃料流量以及 SCR 下游 NO_x 传感器输出值，因此在相关系数计算以及建模中，剔除这三个特征，防止模型通过对公式学习产生虚假的优异表现。

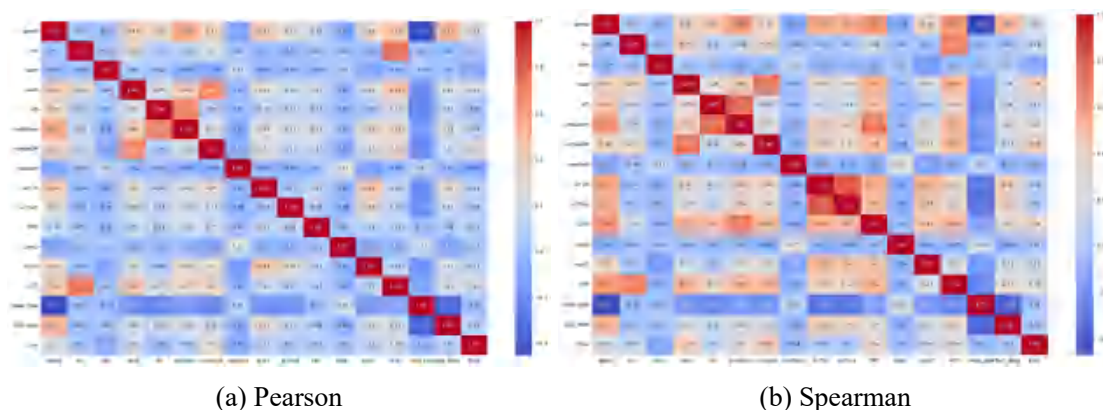


图 4.1 Pearson 和 Spearman 相关系数

根据相关系数可知， NO_x 瞬态排放率与车速（speed）、发动机净输出扭矩（neot）、比功率（VSP）等具有一定的单调相关性。另外，为了考虑环境因素的

影响,将经纬度以及快速路标签(fast_label)加入特征集。因此,初步选用包含11个特征的特征集 Features1 作为基础 NO_x 瞬态排放预测的输入特征。

$$\text{Features1} = \{\text{speed}, \text{Acc}, \text{neot}, \text{eft}, \text{VSP}, \text{sctTin}, \text{sctTout}, \text{DPF}, \text{lng}, \text{lat}, \text{fast_label}\}. \quad (4.1)$$

4.2 基于 XGBoost 的基础 NO_x 排放模型

针对本实验所使用的数据集中存在非线性关联以及多维特征交互的特性,本次实验采用集成学习中的梯度提升决策树(GBDT)框架中的 XGBoost 算法构建 NO_x 排放预测模型。作为高效的梯度提升框架, XGBoost 的三个能力特点:通过决策树组合逼近任意非线性函数的非线性表征能力、高计算效率、内置正则化机制抑制过拟合保证鲁棒性,与本次实验的数据集所需模型的特征较为匹配。相较于传统的单一学习, XGBoost 通过集成策略(Boosting)显著提升模型泛化能力。此外, XGBoost 算法可利用 SHAP 值对特征重要性进行评估,可以为特征消融提供量化依据。因此,选用 XGBoost 算法既满足了复杂排放机制建模需求,又为模型可解释性提供了技术支撑。

以特征集 Features1 为输入特征,在时空数据集上建立基于 XGBoost 算法的 NO_x 排放模型。将模型预测的 NO_x 排放率与原始数据对比,表现如图4.2(a)所示,其决定系数 $R^2 = 0.85$, 均方误差 $\text{MSE} = 0.029$ 。虽然可以解释原始数据集 85% 的数据波动,但在 NO_x 排放较低时,预测效果不好,图中较为离散。

结合图4.2(b)中的平均绝对 SHAP 值分布可知,快速路标签 fast_label 几乎对 NO_x 排放预测没有贡献。另外,针对平均绝对 SHAP 值排名的后五位进行特征消融实验,结果如表4.1所示。根据 ΔR^2 与 ΔRMSE 分析可知,除了 SCR 入口温度(sctTin)与摩擦扭矩(eft)以外,其余三个特征对 NO_x 排放预测结果没有显著的正向影响。

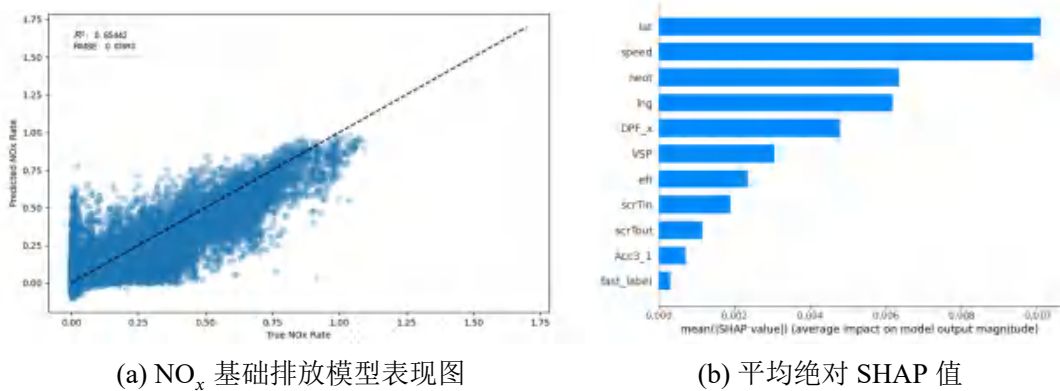


图 4.2 基础 NO_x 排放模型表现

表 4.1 特征消融对比结果

移除特征	ΔR^2	$\Delta RMSE$
eft	-0.006573	0.000831
scrTin	-0.000044	0.000006
scrTout	0.000056	-0.000009
Acc	0.001656	-0.000214
fast_label	0.000506	-0.000065

综合上述分析，从特征集 Features1 中剔除这 Acc、scrTout 以及 fast_label 三个特征，构成包含 8 个特征的新特征集 Features2（其中的 DPF_x 与 DPF 代表同一个特征），用于最终的 NO_x 瞬态排放预测模型的建立。

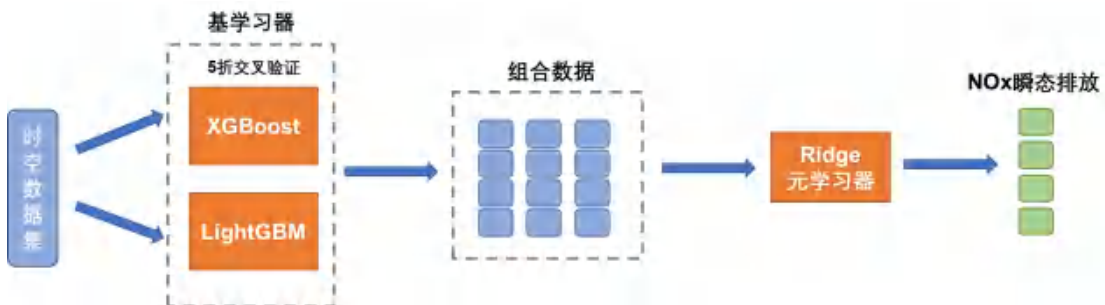
$$\text{Features2} = \{\text{speed, neot, eft, VSP, scrTin, DPF}_x, \text{lng, lat}\}. \quad (4.2)$$

4.3 基于堆叠泛化的 SLXGB-NO_x 排放预测模型

为了提高重型柴油货车 NO_x 排放预测模型的训练效率和预测精度，本研究构建了基于堆叠泛化（Stacking）的混合集成模型 SLXGB。该模型框架通过融合 LightGBM 与 XGBoost 的算法优势，充分挖掘时空数据集中的非线性特征与多维交互关系，以岭回归作为集成策略，对基学习器的预测结果进行加权融合，得出最终预测结果。

4.3.1 模型建立步骤

图4.3为 SLXGB-NO_x 排放预测模型构成示意图。该架构通过分层学习机制，实现了对重型柴油货车 NO_x 瞬态排放的高效建模与精准预测。SLXGB-NO_x 排放预测模型的具体建立步骤如下：

图 4.3 SLXGB-NO_x 排放预测模型

（1）训练集与测试集划分：以 Features2 中的 8 个关键特征构成特征空间 $\mathbf{X} \in R^{n \times 8}$ ，NO_x 瞬态排放率为目标变量 $\mathbf{y} \in R^n$ ，采用时间序列分割法按照 8:2 的比例划分数据集，确保测试集数据时序位于训练集之后，防止未来信息泄露。详

细公式如下：

$$\begin{cases} \mathbf{X}_{\text{train}} = \{\mathbf{x}_i\}_{i=1}^{0.8n}, \mathbf{y}_{\text{train}} = \{\mathbf{y}_i\}_{i=1}^{0.8n}, \\ \mathbf{X}_{\text{test}} = \{\mathbf{x}_i\}_{i=0.8n+1}^n, \mathbf{y}_{\text{test}} = \{\mathbf{y}_i\}_{i=0.8n+1}^n. \end{cases} \quad (4.3)$$

(2) LightGBM 基准模型构建：基于直方图算法的高效特性，设置叶节点数（num_leaves）、学习率（learning_rate）等 8 个核心参数（参数空间见表4.2）。采用贝叶斯优化进行 50 次迭代搜索，通过 5 折交叉验证（n_splits=5）筛选最优参数组合，以最小化验证集均方根误差（RMSE）为目标函数，具体公式如 (4.4) 所示，其中 $\text{RMSE}(\mathbf{y}_{\text{val}}^{(k)}, \hat{\mathbf{y}}_{\text{val}}^{(k)})$ 为第 k 折交叉验证集的 RMSE。

$$\theta_{\text{LGB}}^* = \arg \min_{\theta \in \Theta} \frac{1}{5} \sum_{k=1}^5 \text{RMSE}(\mathbf{y}_{\text{val}}^{(k)}, \hat{\mathbf{y}}_{\text{val}}^{(k)}). \quad (4.4)$$

表 4.2 LightGBM 超参数取值范围

超参数	取值范围	最佳取值
learning_rate	uniform(0.01,0.3)	0.210
max_depth	randint(3,12)	10
n_estimators	randint(100,1000)	825
num_leaves	randint(20,100)	54
subsample	uniform(0.4,0.6)	0.58
colsample_bytree	uniform(0.4,0.6)	0.53
lambda_l1	uniform(0,1)	0.120
lambda_l2	uniform(0,1)	0.472

(3) XGBoost 基准模型优化：针对 XGBoost 的精确贪心算法特性，基于上方建立的 XGBoost 基准模型，设定最大树深度（max_depth）、L2 正则项（reg_lambda）等 9 个超参数，同样通过贝叶斯优化，并结合 5 折交叉验证评估模型的泛化性能，获取最佳参数取值（参数范围及最佳取值见表4.3）。

(4) Stacking 集成架构：建立双层融合框架，第一层基学习器由优化后的 LightGBM 和 XGBoost 构成，第二层采用岭回归作为元学习器。通过 5 折交叉验证，建立最终的 NO_x 瞬态排放预测模型，表达式为 (4.5)。式子中， α 和 β 为基模型权重，取值范围为 [0,1]，且满足 $\alpha + \beta = 1$ ，控制两个树模型（LightGBM 和 XGBoost）的贡献； $f_{\text{LGB}}(\mathbf{x})$ 、 $f_{\text{XGB}}(\mathbf{x})$ 则分别为 LightGBM 与 XGBoost 基学习器的预测结果； γ 为元模型缩放因子， $\mathbf{w}$ 为岭回归对基学习器输出结果的权重向量， b 为偏置项。

表 4.3 XGBoost 超参数范围

超参数	取值范围	最佳取值
learning_rate	uniform(0.01,0.3)	0.135
max_depth	randint(3,12)	10
min_child_weight	randint(1,10)	5
subsample	uniform(0.6,1.0)	0.967
colsample_bytree	uniform(0.6,1.0)	0.813
n_estimators	randint(100,1000)	493
reg_lambda	uniform(0,1)	0.545
reg_alpha	uniform(0,1)	0.432
gamma	uniform(0,1)	0.038

$$f_{\text{stack}}(\mathbf{x}) = \alpha f_{\text{LGB}}(\mathbf{x}) + \beta f_{\text{XGB}}(\mathbf{x}) + \gamma(\mathbf{w}^T[f_{\text{LGB}}(\mathbf{x}); f_{\text{XGB}}(\mathbf{x})] + b). \quad (4.5)$$

通过以上步骤，集成 LightGBM (LGB) 与 XGBoost (XGB) 算法，建立重型柴油货车的 NO_x 瞬态排放预测模型，即 SLXGB-NO_x 排放模型。

4.3.2 模型结果分析

1. 模型特征分析

在 SLXGB-NO_x 瞬态排放模型中，基模型权重如表4.4所示。通过岭回归元学习器的系数权重分析发现，LightGBM 与 XGBoost 在 Stacking 框架中的相对贡献度比为 57:43。与 XGBoost 基准模型对比，LightGBM 的加入以及超参数优化方法的改进使模型的训练效率提升了 10% 左右。

表 4.4 基模型权重

基模型	权重 (α/β)
LightGBM	0.57
XGBoost	0.43

另外，图4.4展示了模型输入特征的 SHAP 值，利用 XGBoost 与 LightGBM 的特征重要性可解释性，对各输入特征进行分析综合分析可知：XGBoost 中，车速 (speed) 的平均绝对 SHAP 值柱状长度最长，是对 NO_x 瞬态排放预测影响最大的特征，表明车速与 NO_x 排放之间存在较强的关联性，这是因为车速的变化会影响发动机的工况，进而影响发动机中燃料燃烧状态，改变 NO_x 的排放情况。另外，发动机净输出扭矩 (neot) 与摩擦扭矩 (eft) 的 SHAP 值次之，也说明了

发动机工作状态对 NO_x 瞬态排放的显著影响。而在 LightGBM 中，可以看出车辆经纬度对 NO_x 瞬态排放有较大的影响，SCR 系统温度（scrTin）与 DPF 压差（DPF_x）的影响次之，说明了位置信息以及尾气处理系统对 NO_x 瞬态排放的影响。

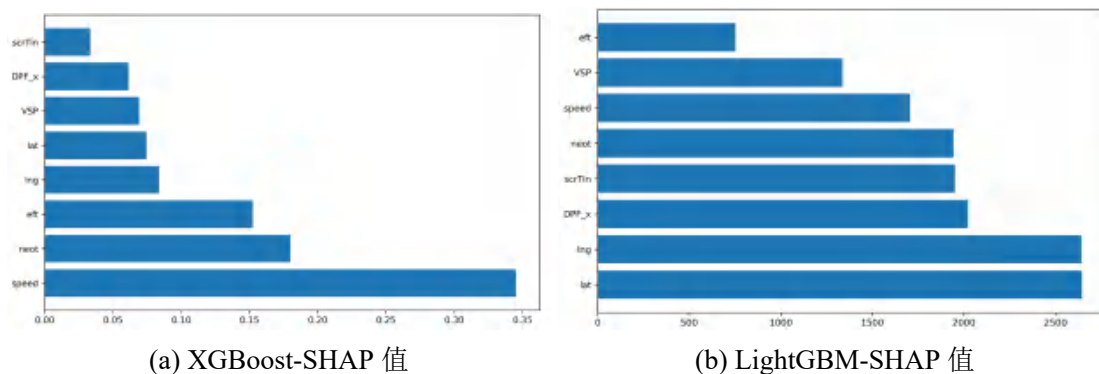


图 4.4 SLXGB-NO_x 排放模型特征 SHAP 值

综合两个基模型对各输入特征重要性的解释可以看出，XGBoost 强于表征瞬态燃烧过程对 NO_x 瞬态排放的影响，而 LightGBM 擅长捕捉后处理动态响应以及环境因素对排放的影响。因此，通过 Stacking 的结合后，二者的输出具有较好的互补性，显著提升了预测的精度与效率。

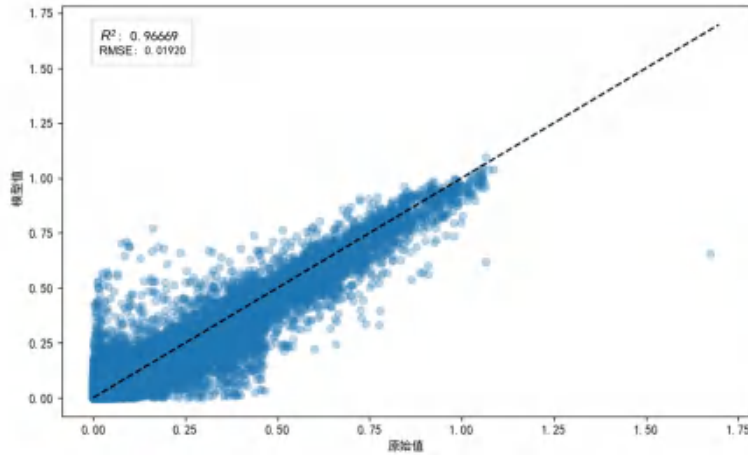
2. 模型结果分析

图4.5直观展示了 SLXGB-NO_x 瞬态排放预测模型的原始-预测值散点分布，图中央的对角斜虚线代表理论上的完全拟合线（ $y=x$ ）。从散点图可以看到，原始-预测数据点都集中且均匀地分布在完全拟合线的两侧，这体现了模型预测值与原始 NO_x 排放率值之间的一致性，表示模型的拟合效果良好。另外，模型的决定系数 $R^2=0.967$ ，表明模型可解释 96.7% 的 NO_x 瞬态排放方差，相较于利用 XGBoost 建立的基准模型提高了 0.113，同时， $RMSE=0.0192$ ，相较于基准降低了 0.00973，模型的预测精准性得到了显著提升。

4.3.3 模型对比

在建立 SLXGB-NO_x 瞬态排放预测模型之前，对多种基准模型进行了测试，以真实-预测值散点图展示模型效果，并利用 R^2 与 RMSE 评价模型的表现，XGBoost 表现见图4.2，其余基础模型表现如图4.6所示。

(1) 多元线性回归模型：相较于一元线性回归，多元线性回归考虑了多个输入参数对目标变量的影响，通过引入多个因素更全面的模拟输入与输出之间的关系，模型表达式为 (4.6)，其中 $\mathbf{x}$ 为输入向量，报个多个输入参数， $\mathbf{k}^T$ 为特征权重， b 则是模型的偏置，目标是通过学习 $\mathbf{k}^T$ 与 b 使输出值 $f(\mathbf{x})$ 尽可能接近目

图 4.5 SLXGB- NO_x 排放模型效果

标变量真实值。

$$f(\mathbf{x}) = \mathbf{k}^T \mathbf{x} + b. \quad (4.6)$$

(2) K 近邻算法 (K-Nearest Neighbor, KNN): 作为一种非参数化模型, KNN 通过计算样本间的欧氏距离, 选择目标点最近的 K 个邻居样本的均值作为预测值。其性能高度依赖于邻域大小 K 与距离度量方式的选择。实验表明 (表4.5), KNN 的 R^2 为 0.654, 显著优于 MLR, 表明 NO_x 排放数据具有局部非线性特征, 但其 RMSE (0.0446) 仍高于树模型, 原因可能在于高维输入下距离计算失效, 且未显式考虑特征间的交互作用。

(3) 反向传播神经网络 (Back Propagation Neural Network, BP): 作为一种误差反向传播算法训练的多层前馈神经网络, BP 神经网络^[47]通过正反向传播不断调整权重与偏置, 并结合梯度下降实现优化, 以最小化误差。以 BP 神经网络为基础, 实现对 NO_x 排放的预测。结果可见4.5, 其中 R^2 为 0.730, RMSE 为 0.0394。虽然相较于线性回归与 KNN 提升, 但不如下方随机森林算法的表现。

(4) 双向长短期记忆网络 (Bi-directional Long Short-Term Memory, BiLSTM): 长短期记忆网络 (Long Short-Term Memory, LSTM)^[48-49]通过引入门控机制, 解决了梯度消失和爆炸问题, 可以捕捉长序列中的长期依赖关系。BiLSTM 则是由正反两个 LSTM 组成, 捕捉双向依赖信息。以 BiLSTM 为基础的 NO_x 预测在 R^2 上的表现为 0.712, RMSE 为 0.0407。相较于 BP 神经网络有所降低, 且在数据量大时, 效率较慢。

(5) 随机森林算法 (RF): 基于 Bagging 集成框架, RF 通过并行训练多棵决策树并聚合结果以提升泛化能力。每棵树仅使用随机子样本及部分特征进行训练, 通过特征扰动降低模型方差。如表4.5所示, RF 的 R^2 达到 0.835, RMSE 降至 0.0308, 验证了集成学习对复杂非线性关系的捕捉能力。但与 XGBoost 相比,

其 R^2 低了 0.019，可能源于两方面：①单棵树的生长策略未引入梯度优化，局部拟合效率较低；②特征随机性虽抑制过拟合，但也损失了部分有效特征组合信息。另外，在高维度预测模型训练中，RF 效率会随着输入参数的增多明显降低，整体效率偏低。

（6）XGBoost 与 LightGBM：基于梯度提升框架，XGBoost 通过二阶导数优化目标函数，并引入正则化项控制模型复杂度，其 R^2 提升至 0.854。LightGBM 则采用单边梯度采样与特征捆绑技术，进一步将 RMSE 降低至 0.0246，验证了高效特征工程与计算优化对预测精度的促进作用。

（7）SLXGB 模型：本文提出的 SLXGB 模型利用 Stacking 集成 LightGBM 与 XGBoost 框架，根据上方的模型结果分析可知，SLXGB 模型的优势源于 LightGBM 与 XGBoost 的算法互补性：XGBoost 通过精确贪心算法和正则化项优化，擅长捕捉发动机瞬态工况的非线性特征；LightGBM 基于直方图优化和特征捆绑技术，能够高效处理高维空间中的环境参数。Stacking 框架通过岭回归融合二者的输出，不仅缓解了单一模型的过拟合风险，还通过元学习器动态调整基模型权重，进一步增强了模型对复杂数据分布的适应性。如表4.5所示，其 R^2 达到 0.967，RMSE 较纯 XGBoost 模型下降 33.6%，表明了集成模型的改进能显著提升模型对瞬态排放特征的解析能力。

表 4.5 模型对比

模型	R^2	RMSE
多元线性回归	0.144	0.0702
KNN	0.654	0.0446
BP	0.730	0.0394
BiLSTM	0.712	0.0407
RF	0.835	0.0308
XGBoost	0.854	0.0289
LightGBM	0.895	0.0246
SLXGB	0.967	0.0192

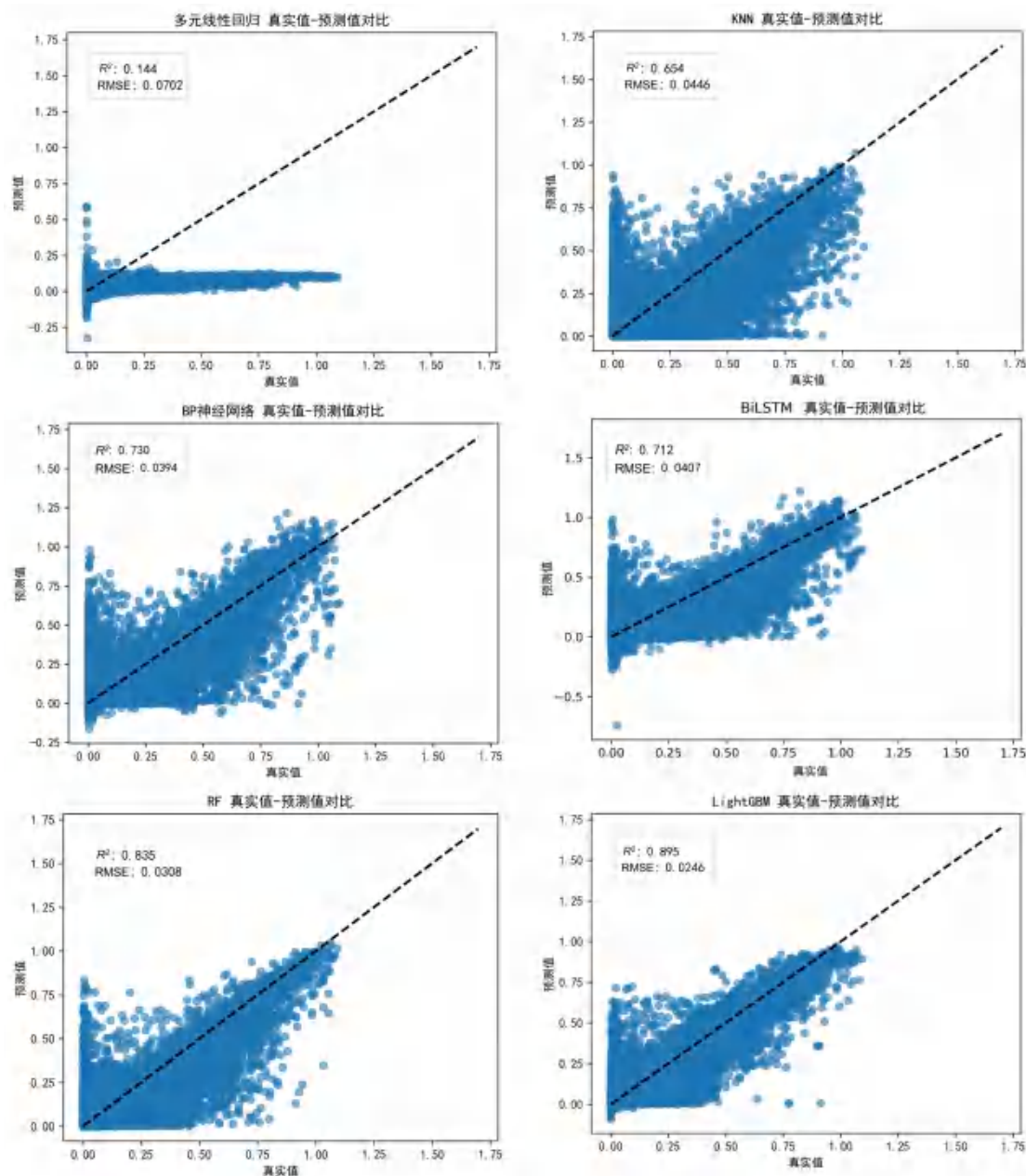


图 4.6 基础模型对比

4.4 成都市高分辨率 NO_x 排放清单

将 NO_x 瞬态排放预测模型用于建立重型柴油货车高分辨率的 NO_x 排放清单是本次实验的最终目的。高分辨率的 NO_x 排放清单可以实现对重型柴油货车的 NO_x 排放定量描述, 评估排放的影响, 预测污染形势可能的走向, 这在重型柴油货车的排放管理工作中有着至关重要的作用, 可以为制定相关减排策略提供关键参考, 同时也为污染管理等政策制订提供数据依据, 促进经济的可持续发展。本研究利用上文建立的 SLXGB- NO_x 瞬态排放预测模型, 结合成都市重型柴油货车逐秒的 OBD 数据, 获取 NO_x 瞬态排放, 并结合 OSM 获取成都市路网地图, 使用 QGIS 进行网格划分, 建立 $1\text{km} \times 1\text{km}$ 的逐小时国六重型柴油货车 NO_x

排放清单，并分析成都市重型柴油货车 NO_x 排放的时空分布特征。具体建立过程如下方内容。

4.4.1 逐小时车辆运行与 NO_x 排放数据

1. 逐小时车辆运行数据集

成都市位于东经 $102.9^\circ \sim 104.8833^\circ$ ，北纬 $30.0833^\circ \sim 31.4333^\circ$ 之间。利用远程监控平台的OBD数据集和GPS定位，根据成都市的经纬度边框，逐小时筛选出成都市范围内24小时的重型柴油货车OBD数据集，并依照第三章3.1中OBD数据筛选与清洗规则获取有效数据集，按照MOVES模型的VSP计算方式，获取VSP工况数据。另外，根据SLXGB- NO_x 模型对输入特征的要求，即Features2，从数据集中提取模型所用的输入数据集。以此获取成都市重型柴油货车每小时的车辆运行数据集，共获取24个数据集。

2. 逐小时车辆运行里程

对于逐秒的数据，其当前速度 v （单位： km/h ）转换为 v_s （ km/s ）即可获得当前的运行里程。然后，将每小时数据中的 v_s 累加，即可获得成都市重型柴油货车每小时的累积行驶里程，统计后如图4.7所示。可以看出，在成都市内，国六重型柴油货车的运行里程从凌晨1点到上午8点，一直呈上升趋势，在8点到下午14点，是车辆运行的高峰时段，并在14点达到逐小时里程数据的最高峰，之后呈现一直下降趋势。

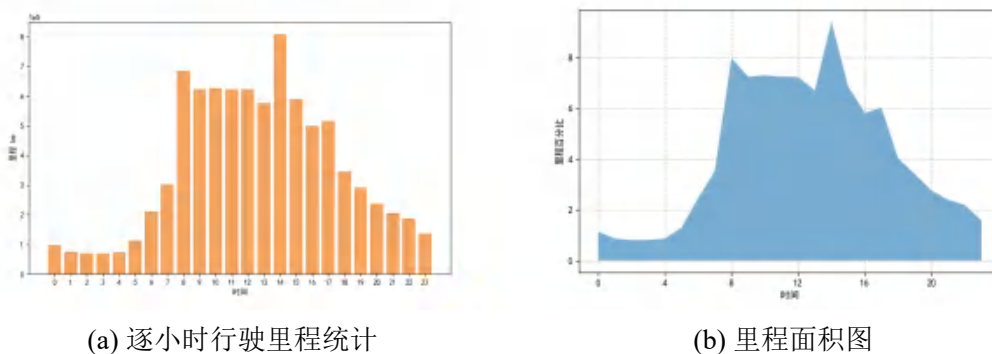


图 4.7 成都市逐小时车辆数据统计

3. 逐小时 NO_x 排放量

基于上方获取的24个车辆运行数据集，输入SLXGB- NO_x 模型后，获取车辆的 NO_x 瞬态排放率（ Enox ），也是车辆的逐秒排放量。然后，将每个数据集中的 Enox 列的数据累加，即可获得每小时的总排放量。统计逐小时的排放量以及变化趋势，获得如图4.8所示，全天的 NO_x 排放呈现出明显的“双峰”分布，这与重型柴油货车逐小时的运行里程变化一致。

累计加和可知，成都市24小时的重型柴油货车 NO_x 排放量为4.2吨。其中，

上午 9 点与下午 14 点逐小时的 NO_x 排放量最高, 分别约为 345.57kg、371.64kg。具体分析可知, 一天之内, 成都市的国六重型柴油货车在凌晨 1 点到 4 点的 NO_x 排放量最低, 之后便呈现快速上升趋势。在交通高峰期上午 8 点至 11 点, NO_x 排放量也处于高峰期。在下午 13 点到 15 点, 是一天之中的第二个 NO_x 排放量高峰期。另外, 统计可知白天 7 点到 18 点时段的重型柴油货车 NO_x 排放占全天排放的 79.9%, 这与重型柴油货车货物运输的工作性质以及工作时间要求极为相关。

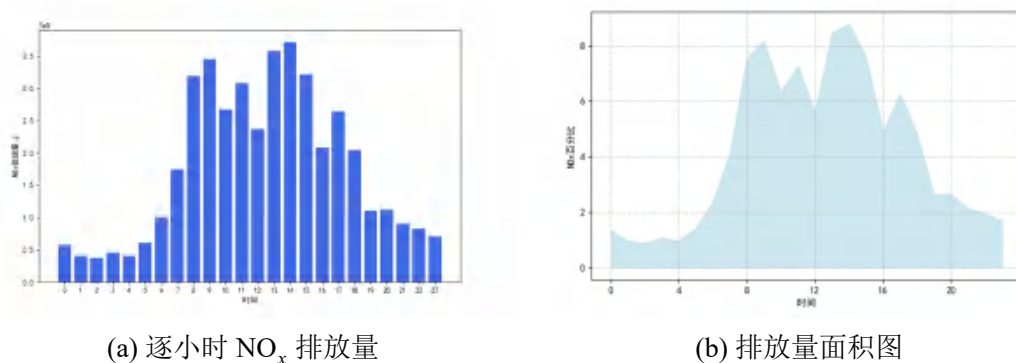


图 4.8 成都市重型柴油货车逐小时 NO_x 排放量

4.4.2 NO_x 网格化排放清单

1. 成都市地图网格化

利用 OSM 获取成都市的路网矢量图, 在 WGS-84 (World Geodetic System 1984) 坐标系下, 使用 QGIS 软件, 利用经纬度将成都市划分为 38556 个 1km×1km 的网格, 并使用数字作为网格的横纵坐标, 对每个网格进行坐标编号, 作为索引, 方便定位到网格。网格划分的具体计算如公式 (4.7), 其中 X 、 Y 分别代表划分网格后的网格的横、纵坐标, 向下取整, x_0 、 y_0 则分别代表所在的经度、纬度。

$$\begin{cases} X = \frac{x_0 - 70}{0.008333}, \\ Y = \frac{y_0 - 10}{0.008333}. \end{cases} \quad (4.7)$$

2. 车辆里程与 NO_x 排放网格化排放清单

(1) 将每小时数据集中的每条记录, 使用经纬度计算对应的网格坐标 (X, Y), 并将其作为网格 ID 加入数据集。

(2) 根据网格 ID 这一标签, 将每小时数据进行分组, 提取每组的 v_s 和 E_{nox} 。通过将每组的 v_s 、 E_{nox} 累加, 即可获得每小时内各个网格的车辆运行里程 ($M_{grid,h}$) 以及 NO_x 排放量 ($E_{grid,h}$), 并按照 (网格 ID, 经度, 纬度, 里程, NO_x 排放量) 的格式, 保存为 CSV 文件。

$$M_{grid,h} = \sum_{i \in grid} (v_s)_{(i,h)}$$

$$E_{grid,h} = \sum_{i \in grid} Enox_{(i,h)}$$

(3) 利用 Python 的 geopandas，结合成都市的矢量路网图，分别绘制每小时的里程和 NO_x 排放网格化的空间分布。如图4.10、4.11以上午 9 时与下午 14 时两个高峰时段为例展示空间分布热力图。

(4) 参考已有研究中成都市的主要道路分布图可知^[50]，对于图4.9中较为明显的环状，从外向内依次为第二绕城高速、绕城高速、三环及二环。结合网格化的里程分布可知，空间上，上午 9 点，成都市重型柴油货车运行集中在成都市的绕城高速和第二绕城高速上，运行集中强度呈现从中心向外，先增强后逐渐降低的趋势。另外，在环状分布以外，成都市周边的一些主要道路上，车辆运行也较为集中。而在下午 14 点，重型柴油货车的运行向城中进一步集中。这是因为重型柴油货车主要进行运输工作，其运行活动与交通干线分布和物流通道相关，因此会在呈现出在主要道路上的集中分布。对比观察 NO_x 排放量的空间分布热力图可知，重型柴油货车 NO_x 的高排放区域也围绕绕城高速聚集，其变化特征与运行里程的空间分布特征一致。二者都呈现出环状聚集以及交通干线聚集的分布特征。

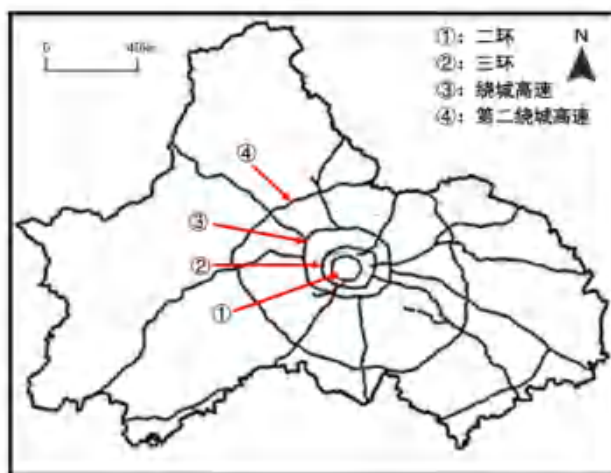


图 4.9 成都市主要道路

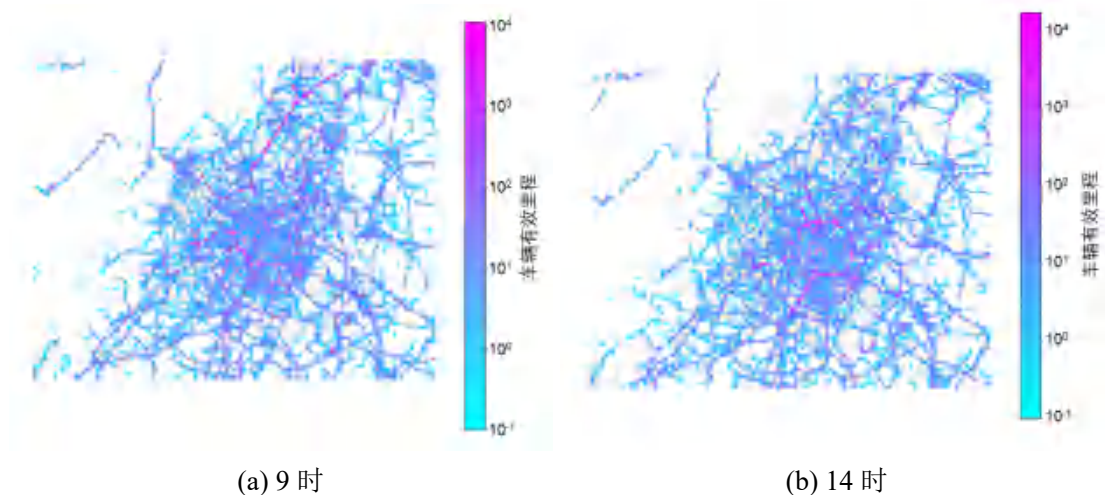
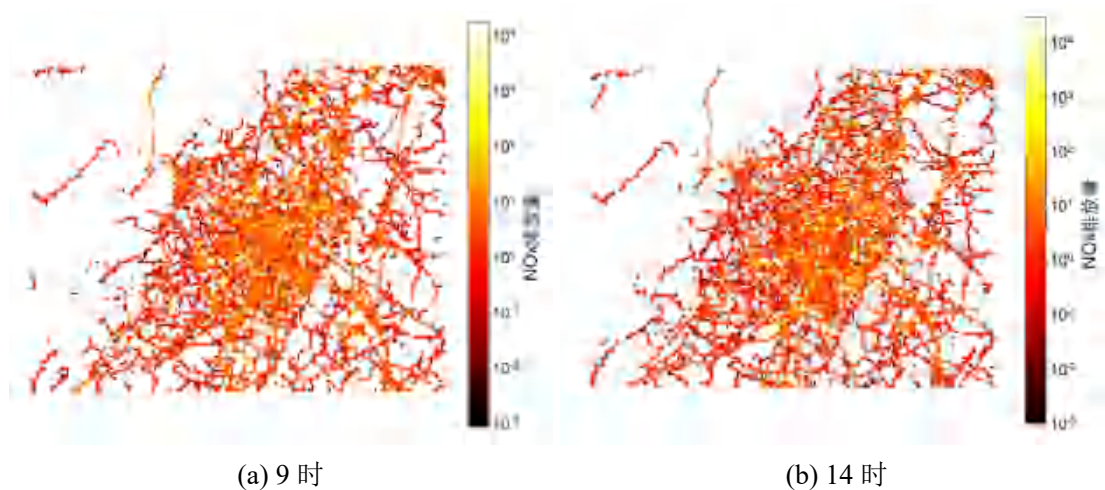


图 4.10 成都市 (a)9 时 (b)14 时里程数据分布

图 4.11 成都市 (a)9 时 (b)14 时 NO_x 排放量分布

4.5 本章小结

本章为了突破传统模型无法计算 NO_x 近实时瞬态排放的缺点，以及相较于计算公式，能够更全面的考虑车辆运行特征对 NO_x 瞬态排放的影响，基于第三章建立的重型柴油货车的时空数据集，以原始 NO_x 排放率为参考值，建立预测重型柴油货车 NO_x 瞬态排放的模型。首先是利用相关系数分析车辆运行参数以及环境特征与 NO_x 瞬态排放之间的相关性，初步获取了 11 个参数作为特征集 Features1。

接着，选择使用 XGBosot 快速建立一个基础 NO_x 瞬态排放预测模型，并利用 SHAP 值展示输入特征对 NO_x 瞬态排放影响的重要程度。结合特征消融实验，剔除对模型无正向作用的参数，获取对 NO_x 瞬态排放最具影响的 8 个参数，构成新的特征集 Features2。

然后，以优化后的 LightGBM 和 XGBoost 构成基学习器，使用岭回归作为

元学习器，获取了 Stacking 集合架构的重型柴油货车 NO_x 瞬态排放预测模型，即 SLXGB- NO_x 排放模型。相较于 XGBoost 基础模型， R^2 达到 0.967，提升了约 13%，RMSE 降低了 33%。且在与其他基准模型相比，SLXGB 模型不仅在预测精度上得到了明显提升，相较于其他的树模型，预测效率也更高效。

最后，基于成都市远程监控平台的重型柴油货车 OBD 数据，利用已构建的 SLXGB- NO_x 瞬态排放预测模型，获取了成都市重型柴油货车的 NO_x 瞬态排放，并通过累加，计算成都市重型柴油货车逐小时的 NO_x 排放量。另外，利用 QGIS 网格化成都市的路网地图，结合 Python 展示网格化排放清单热力图。结合逐小时的成都市重型柴油货车 NO_x 排放可知：全天 NO_x 排放日间占比高达 79.9%，且有明显的双峰值。结合网格化清单可知：重型柴油货车的 NO_x 排放围绕成都市绕城高度聚集，呈现环状分布。

第5章 总结与展望

5.1 本文工作总结

在我国经济持续高速发展的大背景下，国内物流运输的需求呈爆发式增长，重型柴油货车作为公路货运的核心运力，其保有量也随之快速攀升。但与此同时，这一增长态势背后的环境代价逐渐显现，重型柴油货车尾气排放已成为城市大气污染的重要贡献源，其中氮氧化物（ NO_x ）的过量排放问题尤为突出，对空气质量和公众健康构成严峻威胁。精准掌握重型柴油货车 NO_x 排放状况，对其进行准确的自下而上核算，构建准确的 NO_x 排放清单，不仅是解决当前大气污染难题的关键举措，更是为科学制定大气污染治理政策、高效开展环境管理工作提供坚实数据支撑与决策依据的迫切需求。

建立宏观层面准确的 NO_x 排放清单需要可靠的重型柴油货车 NO_x 排放因子，而中微观层面高分辨率的 NO_x 排放清单的建立则需要通过排放模型获取车辆的瞬态排放数据。对于 NO_x 排放因子，当前研究中多基于 PEMS 或实验室测试采集的数据来获取，但此类方法存在显著局限性：PEMS 测试数据获取过程繁琐，成本高且样本量有限，实验室测试难以反映真实道路运行工况，导致数据不确定性较大。对于排放模型，传统排放模型多基于统计或者历史研究数据建立排放清单，无法捕捉车辆近实时瞬态排放特征，时空分辨率和精度不足。

为了解决准确的 NO_x 排放因子获取以及高分辨率 NO_x 排放清单建立过程中的两个关键问题，本文的主要工作内容与结论如下：

（1）针对基于 PEMS 或者实验数据样本量有限、工况代表性不足，导致获取的重型柴油货车 NO_x 排放因子精度低、反映不全面的问题，本文基于 OBD 数据与道路信息，建立耦合驾驶工况-道路特征的时空数据集，在涵盖范围与数据量上得到显著扩充，并通过该数据集计算获取国六重型柴油货车的 NO_x 排放因子，填补了真实道路下的国六重型柴油货车 NO_x 排放因子的空白。具体过程为：一，利用 OBD 设备采集重型柴油货车的近实时运行数据，并在该数据集的处理过程中，提出一种分阶段插值策略，结合线性插值、多层感知机以及随机森林算法，实现对 OBD 数据的插值处理，获取完整可用数据集。二，结合 MOVES 模型中的 VSP 计算方法，并根据重型柴油货车的质量特征调整相关参数，获取车辆的 VSP 工况数据，并按照 VSP 区间聚类划分，以编号 VSP Bin 代表工况特征。三，结合 GADM 与 OSM 获取路网地图，通过逐级定位获取车辆所在运行道路，提取道路信息，进一步按照道路类型将道路分为快速路、主干路、次支路三大类，并以此为基准进行车辆行程划分，获取车辆道路特征。通过以上三个步骤以及数据融合处理，获取国六重型柴油货车的近实时时空数据集，基于该数据

集,分析车辆 VSP 工况- NO_x 排放变化特征并计算 NO_x 排放因子。其中,国六重型柴油货车真实道路下的基准排放因子为 1.64g/km ,通过与国六排放标准对重型柴油货车 NO_x 排放限值的要求以及《道路机动车大气污染物排放清单编制技术指南》中的国五排放因子对比可知,该基准排放因子符合实际。另外,快速路的 NO_x 排放因子为 1.82g/km ,较主干路 (1.28g/km) 与次支路 (1.24g/km) 高出 29.7%-31.9%,这是由于不同道路上,车辆的运行特征不同对 NO_x 排放造成的差异化影响。基于时空数据集获取的重型柴油货车 NO_x 排放因子综合考虑了车辆工况与道路特征对 NO_x 排放的影响,相较于统计数据更加准确且考虑了实际环境影响。

(2) 针对传统模型无法捕捉重型柴油货车 NO_x 近实时排放动态特征的缺陷,本文基于已构建的时空数据集,提出了一个重型柴油货车 NO_x 瞬态排放预测模型,即 SLXGB- NO_x 排放模型。该模型一方面利用 LightGBM 捕捉重型柴油货车运行过程中,车辆经纬度以及 SCR 系统等因素对 NO_x 瞬态排放的影响,一方面利用 XGBoost 捕捉车辆自身运行状态对 NO_x 排放的影响。通过 LightGBM 和 XGBoost 结合构成基学习器,岭回归作为元学习器,利用 Stacking 建立双层架构的集成模型,通过该模型实现对重型柴油货车 NO_x 瞬态排放的预测。与其他基准模型相比,该模型在预测准确性与时间效率上都表现出良好的竞争力。最后,基于成都市重型柴油货车的 OBD 数据,利用该模型建立了成都市 $1\text{km}\times 1\text{km}$ 的逐小时重型柴油货车 NO_x 排放清单,并利用网格化地图进行清单可视化,分析可知成都市的国六重型柴油货车的 NO_x 排在时间上呈现出明显的“双峰”分布,其中日间 NO_x 排放量占全体的 79.9%,在空间上沿绕城高速形成环状聚集带,该特征与已有研究中的成都市排放分布特征趋势一致。

5.2 未来展望

虽然本次研究基于时空数据集获取了重型柴油货车真实道路下的 NO_x 排放因子,构建了 NO_x 瞬态排放预测模型,能够为相关政策的制订提供精准数据支持,为后续相关研究在考虑真实环境因素和瞬态排放获取方面提供新思路与参考,但仍在研究对象以及对排放的影响因素的相关方面存在一定的不足:

(1) 环境因素量化分析不足:为了更加全面的研究外部环境对 NO_x 排放的影响,有必要加入环境温度、湿度、车辆负载变化等因素。这些因素的变化会对车辆的排放直接或间接的产生影响。因此,未来可以着重从加强环境因素与污染物排放的关联建模下手,将更多的实际环境数据考虑到排放影响因素中,并进一步综合研究这些复杂环境对 NO_x 排放的综合影响。

(2) 模型跨域泛化与部署:移动源种类繁多,不同类型之间的排放特征差异

较大，当前模型尚未针对不同排放标准（如国四、国五车辆）或混合动力车型进行跨标准、跨类型的迁移性测试，泛化能力需要进一步加强。因此，未来可以通过多方渠道，获取不同排放标准的车辆的实际运行数据，来用于泛化该模型，实现模型功能的进一步完善，并通过优化模型的算法，提升模型效率，尝试开发边缘计算模块，实现模型的轻量化部署。

参考文献

- [1] 中华人民共和国生态环境部. 中国移动源环境管理年报(2023年)[EB/OL]. 2023. https://www.mee.gov.cn/hjzl/sthjzk/ydyhjl/202312/t20231207_1058460.shtml.
- [2] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Diesel and gasoline engine exhausts and some nitroarenes[M]. Lyon: International Agency for Research on Cancer, 2012.
- [3] 杜强. 机动车尾气排放对城市大气环境的影响与控制[J]. 资源节约与环保, 2023(9): 114-117.
- [4] 重型柴油车污染物排放限值及测量方法(中国第六阶段): GB 17691-2018[S]. 北京: 中国国家标准化管理委员会, 2018.
- [5] 王燕军, 王鸣宇, 等. 国外机动车排放模型综述研究[J]. 环境与可持续发展, 2020, 45(5): 6.
- [6] U.S. Environmental Protection Agency. Motor vehicle emission simulator (MOVES)[EB/OL]. 2025. <https://www.epa.gov/moves/latest-version-motor-vehicle-emission-simulator-moves>.
- [7] Emisia S.A. COPERT | emission calculation software[EB/OL]. 2023. <https://copert.emisia.com/>.
- [8] University of California-Riverside, College of Engineering-Center for Environmental Research and Technology. CMEM model[EB/OL]. 2025. <https://www.cert.ucr.edu/cmем/cmем-model>.
- [9] 何巍楠, 刘莹, 孙胜阳, 等. 基于 HBEFA 的城市交通温室气体排放模型—以北京本地化建模为例[J]. 交通运输系统工程与信息, 2014, 14(4): 222.
- [10] 曹杨, 郭园园, 曹罡, 等. MOVES 模型微观层次参数的深圳本土化研究[J]. 交通信息与安全, 2017, 35(2): 9.
- [11] 周华, 赵宏伟, 吴蒙蒙, 等. IVE 模型中区间划分方法对排放因子估算的影响[J]. 中国环境科学, 2019, 39(2): 560-564.
- [12] 宋晓伟, 郝永佩, 朱晓东. 长三角城市群机动车污染物排放清单建立及特征研究[J]. 环境科学学报, 2020, 40(1): 90-101.
- [13] Gao C, Gao C, Song K, et al. Vehicle emissions inventory in high spatial-temporal resolution and emission reduction strategy in Harbin-Changchun megalopolis[J]. Process Safety and Environmental Protection, 2020, 138: 236-245.
- [14] Nagpure A S, Gurjar B R, Kumar P. Impact of altitude on emission rates of ozone precursors from gasoline-driven light-duty commercial vehicles[J]. Atmospheric Environment, 2011, 45(7): 1413-1417.

- [15] Wallace H, Jobson B, Erickson M, et al. Comparison of wintertime CO to NO_x ratios to MOVES and MOBILE6.2 on-road emissions inventories[J]. *Atmospheric Environment*, 2012, 63: 289-297.
- [16] Ritner M, Westerlund K K, Cooper C D, et al. Accounting for acceleration and deceleration emissions in intersection dispersion modeling using MOVES and CAL3QHC[J]. *Journal of the Air and Waste Management Association*, 2013, 63(6): 724-736.
- [17] The International Council on Clean Transportation. Global overview of on-board diagnostic (OBD) systems for heavy-duty vehicles[EB/OL]. 2015. https://theicct.org/sites/default/files/publications/ICCT_Overview_OBD-HDVs_20150209.pdf.
- [18] Wenjun L, Yiming Z, Wenbin L. A wireless and real-time monitoring system design for car networking applications[J]. *International Journal of Antennas and Propagation*, 2013, 2013(1): 298543.
- [19] 王瑞宾. 基于数据驱动的重型柴油车排放模型修正问题研究[D]. 合肥: 安徽大学, 2023.
- [20] Medashe M O, Abolarin M S. Specifications and analysis of digitized diagnostics of automobiles: A case study of on board diagnostic (OBD II)[J]. *International Journal of Engineering Research and Technology*, 2020, 9(1): 91-105.
- [21] Zhang S, Zhao P, He L, et al. On-board monitoring (OBM) for heavy-duty vehicle emissions in China: Regulations, early-stage evaluation and policy recommendations[J]. *Science of the Total Environment*, 2020, 731: 139045.
- [22] Jiang Y, Yang J, Tan Y, et al. Evaluation of emissions benefits of OBD-based repairs for potential application in a heavy-duty vehicle inspection and maintenance program[J]. *Atmospheric Environment*, 2021, 247: 118186.
- [23] Wang X, Qiu Z, Liu Z. Urban road BC emissions of LDGVs: Machine learning models using OBD/PEMS data[J]. *Chemosphere*, 2024, 365: 143348.
- [24] Hao L, Ren Y, Lu W, et al. Assessment of heavy-duty diesel vehicle NO_x and CO₂ emissions based on OBD data[J]. *Atmosphere*, 2023, 14(9): 1417.
- [25] Lv Z, Zhang Y, Ji Z, et al. A real-time NO_x emission inventory from heavy-duty vehicles based on on-board diagnostics big data with acceptable quality in China[J]. *Journal of Cleaner Production*, 2023, 422: 138592.
- [26] Fernandes P, Macedo E, Bahmankhah B, et al. Are internally observable vehicle data good predictors of vehicle emissions[J]. *Transportation Research Part D: Transport and Environment*, 2019, 77: 252-270.
- [27] Rimpas D, Papadakis A, Samarakou M. Obd-ii sensor diagnostics for monitoring vehicle operation and consumption[J]. *Energy Reports*, 2020, 6: 55-63.
- [28] Rosero F, Fonseca N, López J M, et al. Real-world fuel efficiency and emissions from an urban

- diesel bus engine under transient operating conditions[J]. *Applied Energy*, 2020, 261: 114442.
- [29] Pavlovic J, Fontaras G, Broekaert S, et al. How accurately can we measure vehicle fuel consumption in real world operation[J]. *Transportation Research Part D: Transport and Environment*, 2021, 90: 102666.
- [30] Rivera-Campoverde N D, Muñoz-Sanz J L, Arenas-Ramirez B D V. Estimation of pollutant emissions in real driving conditions based on data from OBD and machine learning[J]. *Sensors*, 2021, 21(19): 6344.
- [31] Tolstikhin I, Houlsby N, Kolesnikov A, et al. Mlp-mixer: an all-MLP architecture for vision[C]//*Proceedings of the 35th International Conference on Neural Information Processing Systems: Vol. 34*. Curran Associates, Inc., 2021: 24261-24272.
- [32] Breiman L. Random Forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [33] Ge Z, Li W, Wang J, et al. Real-road NO_x and CO₂ emissions of city and highway China-6 heavy-duty diesel vehicles[J]. *Journal of Environmental Sciences*, 2023, 149: 330-341.
- [34] Chen T, Guestrin C. XGBoost: A scalable tree boosting system[C]//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2016: 785-794.
- [35] Ke G, Meng Q, Finley T, et al. Lightgbm: a highly efficient gradient boosting decision tree[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2017: 3149-3157.
- [36] 重型车排放远程监控技术规范第 1 部分车载终端: HJ 1239.1—2021[S]. 北京: 中华人民共和国生态环境部, 2021.
- [37] Jiménez-Palacios J L. Understanding and quantifying motor vehicle emissions with vehicle specific power and TILDAS remote sensing[D]. Cambridge: Massachusetts Institute of Technology, 1999.
- [38] 城市综合交通体系规划标准: GB/T 51328-2018[S]. 北京: 中华人民共和国住房和城乡建设部, 2018.
- [39] U.S. Environmental Protection Agency. Final facility specific speed correction factors [EB/OL]. 2001. <https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P10022E4.txt>.
- [40] 冯谦, 刘保献, 杨妍妍, 等. 基于远程监测的北京市柴油车实际道路 NO_x 与 CO₂ 排放特征[J]. *中国环境科学*, 2023, 43(8): 4418-4426.
- [41] Wang X, Song G, Wu Y, et al. A NO_x emission model incorporating temperature for heavy-duty diesel vehicles with urea-SCR systems based on field operating modes[J]. *Atmosphere*, 2019, 10(6): 337.
- [42] Asif M, Zhang Y, Lin W. Urea-SCR temperature investigation for NO_x control of diesel engine [C]//*MATEC Web of Conferences: Vol. 26*. EDP Sciences, 2015: 03002.

- [43] Xu D, Gao Z, Guo Y, et al. Study on fuel consumption and emission characteristics of China VI heavy duty vehicle based on vehicle specific power[C]//E3S Web of Conferences. VESEP2020: Vol. 268. EDP Sciences, 2021: 01055.
- [44] 宋国华, 吕海欧, 李祖芬, 等. 基于比功率分布的排放因子速度修正误差控制研究[J]. 交通运输系统工程与信息, 2021, 21(6): 272-282.
- [45] 马帅, 黄志辉, 纪亮, 等. 基于车载测试的国 VI 重型柴油车 CO₂ 和 NO_x 排放研究[J]. 环境科学学报, 2022, 42(2): 341-350.
- [46] 中华人民共和国生态环境部. 道路机动车大气污染物排放清单编制技术指南 (试行) [EB/OL]. 2014. <https://www.mee.gov.cn/gkml/hbb/bgth/201407/W020140708387895271474.pdf>.
- [47] Uslu S, Celik M B. Prediction of engine emissions and performance with artificial neural networks in a single cylinder diesel engine using diethyl ether[J]. Engineering Science and Technology, an International Journal, 2018, 21(6): 1194-1201.
- [48] Howlader A M, Patel D, Gammariello R. Data-driven approach for instantaneous vehicle emission predicting using integrated deep neural network[J]. Transportation Research Part D: Transport and Environment, 2023, 116: 103654.
- [49] 邓明星, 欧阳含笑, 钱枫, 等. 基于改进 LSTM 的重型柴油车远程监测 NO_x 浓度缺失数据填补[J]. 环境科学学报, 2023, 43(11): 245-257.
- [50] 潘玉瑾, 李媛, 陈军辉, 等. 基于交通流的成都市高分辨率机动车排放清单建立[J]. 环境科学, 2020, 41(8): 3581-3590.
- [51] Guo D, Wang Z G, Sun L, et al. Study on gasoline vehicle emission inventory considering regional differences in China[J]. Journal of Advanced Transportation, 2018, 2018(1): 7497354.
- [52] Shahbazi H, Taghvace S, Hosseini V, et al. A gis based emission inventory development for Tehran[J]. Urban Climate, 2016, 17: 216-229.
- [53] Zhou Z, Tan Q, Liu H, et al. Emission characteristics and high-resolution spatial and temporal distribution of pollutants from motor vehicles in Chengdu, China[J]. Atmospheric Pollution Research, 2019, 10(3): 749-758.
- [54] Yang D, Zhang S, Niu T, et al. High-resolution mapping of vehicle emissions of atmospheric pollutants based on large-scale, real-world traffic datasets[J]. Atmospheric Chemistry and Physics, 2019, 19(13): 8831-8843.
- [55] Carlos G, Christian V, Frederik D S, et al. From OBD to connected diagnostics: a game changer at fleet, vehicle and component level[C]//IFAC-PapersOnLine: Vol. 54. 2021: 558-563.
- [56] 杨艳, 陈建华, 张鹏, 等. 基于实际行驶里程的重型货车动态排放清单构建方法与应用[J]. 环境科学研究, 2023, 36(10): 1905-1914.
- [57] Wu L, Chang M, Wang X, et al. Development of the real-time on-road emission (ROE v1.0)

- model for street-scale air quality modeling based on dynamic traffic big data[J]. *Geoscientific Model Development*, 2020, 13(1): 23-40.
- [58] Chen B, Hong C, Kan H. Exposures and health outcomes from outdoor air pollutants in China [J]. *Toxicology*, 2004, 198(1): 291-300.
- [59] Lv Z, Zhang Y, Ji Z, et al. A real-time NO_x emission inventory from heavy-duty vehicles based on on-board diagnostics big data with acceptable quality in China[J]. *Journal of Cleaner Production*, 2023, 422: 138592.
- [60] Jiang H, Zhang H, Fu M, et al. Recent advances and perspectives towards emission inventories of mobile sources: Compilation approaches, data acquisition methods, and case studies[J]. *Journal of Environmental Sciences*, 2023, 123: 460-475.
- [61] 葛子豪, 尹航, 徐龙, 等. 基于远程监控的重型车氮氧化物和二氧化碳排放研究[J]. *中国环境科学*: 1-9.
- [62] Wang J, Wang R, Yin H, et al. Assessing heavy-duty vehicles (HDVs) on-road NO_x emission in China from on-board diagnostics (OBD) remote report data[J]. *Science of the Total Environment*, 2022, 846: 157209.
- [63] 柴油车污染物排放限值及测量方法（自由加速法及加载减速法）: GB 3847-2018[S]. 北京: 中华人民共和国生态环境部, 2018.
- [64] 汽油车污染物排放限值及测量方法（双怠速法及简易工况法）: GB 18285-2018[S]. 北京: 中华人民共和国生态环境部, 2018.
- [65] Guo J, Ge Y, Hao L, et al. On-road measurement of regulated pollutants from diesel and CNG buses with urea selective catalytic reduction systems[J]. *Atmospheric Environment*, 2014, 99: 1-9.
- [66] Yang L, Zhang S, Wu Y, et al. Evaluating real-world CO₂ and NO_x emissions for public transit buses using a remote wireless on-board diagnostic (OBD) approach[J]. *Environmental Pollution*, 2016, 218: 453-462.
- [67] Wu Y, Zhang S J, Li M L, et al. The challenge to NO_x emission control for heavy-duty diesel vehicles in China[J]. *Atmospheric Chemistry and Physics*, 2012, 12(19): 9365-9379.
- [68] 李刚, 尹航, 等. 重型车排放远程监控技术规范制订思路及实施建议[J]. *环境工程技术学报*, 2023, 13(2): 867-872.
- [69] 任烁今, 张超, 李刚, 等. 基于远程监控三区移动平均窗口法重型车排放特征[J]. *汽车安全与节能学报*, 2023, 14(3): 375-384.
- [70] Ge Y, Hou P, Lyu T, et al. Machine learning-aided remote monitoring of NO_x emissions from heavy-duty diesel vehicles based on OBD data streams[J]. *Atmosphere*, 2023, 14(4): 651.
- [71] Sun S, Sun L, Liu G, et al. Developing a vehicle emission inventory with high temporal-spatial resolution in Tianjin, China[J]. *Science of the Total Environment*, 2021, 776: 145873.

- [72] Jiang H, Zhang H, Fu M, et al. Recent advances and perspectives towards emission inventories of mobile sources: Compilation approaches, data acquisition methods, and case studies[J]. *Journal of Environmental Sciences*, 2023, 123: 460-475.
- [73] Sun S, Jiang W, Gao W. Vehicle emission trends and spatial distribution in Shandong province, China, from 2000 to 2014[J]. *Atmospheric Environment*, 2016, 147: 190-199.
- [74] Yu Q, Yang Y, Xiong X, et al. Assessing the impact of multi-dimensional driving behaviors on link-level emissions based on a portable emission measurement system (PEMS)[J]. *Atmospheric Pollution Research*, 2021, 12(1): 414-424.
- [75] Stewart G B, Dajnak D, Davison J, et al. New NO_x and NO₂ vehicle emission curves, and their implications for emissions inventories and air pollution modelling[J]. *Urban Climate*, 2024, 57: 102103.

致谢

窗外春风料峭，忽冷忽热，将我的思绪吹成一团乱麻。校园生活就要接近尾声了，此刻，我正在书写致谢，不知道今后是否还有机会再次踏进校园读书，回忆二十余载的求学生涯，酸甜苦辣一应俱全。可惜我才疏学浅，只能用这质朴的文字表达我最真挚的感谢。

师者如舟，带领我前往知识的海洋。我想要感谢我的校内导师赵云波老师，感谢他在从我入学开始就帮我答疑解惑，指导并帮助我完成学业，给予我耐心的教诲。我想要感谢我的实践导师尹航老师、黄志辉老师、何卓识老师，感谢在实践期间，他们对我实习工作的悉心指导，让我学习到了交叉领域的学科知识，拓宽了自己的眼界。非常感恩与各位良师的相遇，让我拥有了一段美好的学习时光。各位良师带给我的温暖都会成为一股支撑我不断前进的动力。

“人之相识，贵在相知，人之相知，贵在知心。”我想要感谢我的朋友们，感谢她们对我无微不至的关心，感谢她们愿意在我迷茫时拉起我的手，带着我向前走，感谢她们愿意一直陪伴在我的身边，哪怕是听我讲一些无聊废话。感恩我的朋友们一直陪在我的身边，时光从未停止流逝，但她们好像为我建立了一堵“时间墙”，让我的青春从未消逝。另外，我也想要感谢我的十位“电子朋友”，从未想过他们会陪伴我将近整个读研时光，是他们告诉“要敢于开口”、“要先去做，再去想”，激励我更勇敢的生活。

春晖寸草，山高海深。感谢父母对我不遗余力的托举，让我拥有看世界的机会与勇气；感谢父母成为我坚强的后盾，支撑我这二十余载的求学生涯；感谢我的弟弟为我的乏味的生活带来了许多欢乐；感谢我每一位家人对我的关怀与照顾。家人的爱与温暖无时无刻不包围着我，是我前行道路上的底气，陪我经历了人生一个又一个的重要阶段，让我积极乐观的面对生活。

致谢来到了结尾，眼底氤氲的水汽又再次模糊了窗外的灯火，校园里的白玉兰不知何时悄然绽放，洁白的花瓣在春风中轻轻摆动，仿佛在向我挥手告别。虽然前方道路未知且漫长，但我依然会带着对生活的热情，积极地迎接每一天！

在读期间取得的科研成果

参与的科研项目：

- (1) 在用机动车高污染成因诊断与治理关键技术研发及示范，国家重点研发计划-项目，2022-2026.
- (2) 时空大数据驱动的移动源近实时排放表征技术研发，国家重点研发计划-课题，2023-2026.